

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

На правах рукопису
УДК 004.942:519.216.3

До захисту допущено
В. о. завідувача кафедри ММСА

О.Л.Тимошук

«__» _____ 2018 р.

Магістерська дисертація

на здобуття ступеня магістра за спеціальністю 122 Комп'ютерні науки
на тему: «Процеси з довгою пам'яттю: моделювання і прогнозування»

Виконав:

студент II курсу, групи КА-71 мп

Шляков Сергій Віталійович

Керівник: професор кафедри ММСА,
д.т.н., професор, Бідюк П.І.

Рецензент: професор ФТІ
д.т.н., професор, Архіпов О.Є.

Засвідчую, що у цій магістерській дисертації
немає запозичень з праць інших авторів
без відповідних посилань
Студент _____

Київ
2018

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

Рівень вищої освіти — другий (магістерський)

Спеціальність (спеціалізація) — 124 «Системний аналіз» («Системний аналіз і управління»)

ЗАТВЕРДЖУЮ

В. о. завідувача кафедри ММСА

О. Л.

Тимошук

«___» _____ 2018 р.

ЗАВДАННЯ

на магістерську дисертацію студенту Шлякову Сергію Віталійовичу

1. Тема дисертації: «Процеси з довгою пам'яттю: моделювання і прогнозування», науковий керівник дисертації Бідюк Петро Іванович, професор, доктор технічних наук, затверджені наказом по університету від «07» листопада 2018 р. № 4121-с

2. Термін подання студентом дисертації: _____

3. Об'єкт дослідження: процеси з довгою пам'яттю.

4. Предмет дослідження: математичні моделі для опису процесів із довгою пам'яттю, методи їх прогнозування на основі статистичних даних.

5. Перелік завдань, які потрібно розробити:

- 1) Огляд технічної літератури за темою роботи;
- 2) Дослідження актуальності обраної теми;
- 3) Вибір методів для моделювання і прогнозування;
- 4) Збір вхідних даних;
- 5) Виконання обчислювальних експериментів;
- 6) Аналіз результатів моделювання і прогнозування;
- 7) Проведення аналізу ринкових можливостей запуску стартап-проекту;
- 8) Підготовка ілюстративного матеріалу;
- 9) Оформлення пояснювальної записки.

6. Орієнтовний перелік графічного (ілюстративного) матеріалу:

- 1) Постановка завдання дослідження;
- 2) Методи інтелектуального аналізу даних;
- 3) Наукова новизна результатів.

7. Орієнтовний перелік публікацій:

8. Дата видачі завдання: _____

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1	Отримання завдання на дипломну роботу	07.09.2018 – 09.09.2018	
2	Огляд технічної літератури за темою	10.09.2018 – 30.09.2018	
3	Дослідження актуальності обраної теми	01.10.2018 – 07.10.2018	
4	Вибір методів для моделювання і прогнозування	08.10.2018 – 14.10.2018	
5	Збір вхідних даних	15.10.2018 – 21.10.2018	
6	Виконання обчислювальних експериментів	22.10.2018 – 28.10.2018	
7	Аналіз результатів моделювання і прогнозування	29.10.2018 – 04.11.2018	
8	Проведення аналізу ринкових можливостей запуску стартап-проекту	05.11.2018 – 11.11.2018	
9	Підготовка ілюстративного матеріалу	12.11.2018 – 18.11.2018	
10	Оформлення пояснювальної записки	19.11.2018 – 26.11.2018	

Студент

Шляков С.В.

Науковий керівник дисертації

Бідюк П.І

РЕФЕРАТ

Магістерська дисертація: 120 с., 14 рис., 23 табл., 1 додаток, 12 джерел.

Об'єкт дослідження – процеси з довгою пам'яттю: моделювання і прогнозування.

Предмет дослідження – математичні моделі для опису процесів із довгою пам'яттю, методи їх прогнозування на основі статистичних даних.

Мета роботи – підвищення якості моделей для процесів із довгою пам'яттю шляхом удосконалення системної методології побудови моделей та комплексного застосування кількох методів інтелектуального аналізу даних.

Методи дослідження – ARFIMA – FIAPARCH, ARFIMA – GARCH, KPSS, ADF.

Актуальність – побудова моделей, що допоможе при дослідженні та прогнозуванні процесів із довгою пам'яттю.

Проведений аналіз отриманих результатів, виконано аналіз отриманої прогностичної моделі.

Шляхи подальшого розвитку предмету дослідження – розвиток методів прогнозування та дослідження процесів із довгою пам'яттю.

ПРОЦЕСИ ІЗ ДОВГОЮ ПАМ'ЯТТЮ, ARFIMA, FIAPARCH, ПОКАЗНИК ХЕРСТА, KPSS, ADF.

ABSTRACT

Master's thesis: 120 p., 14 fig., 23 tabl., 1 appendix, 12 sources.

The object of research – processes with long memory: modeling and forecasting.

Subject of research – mathematical models for the description of processes with long memory, methods of their forecasting on the basis of statistical data.

The purpose of the work is to improve the quality of models for processes with long memory by improving the system methodology of constructing models and the complex application of several methods of data mining.

Research methods - ARFIMA - FIAPARCH, ARFIMA - GARCH, KPSS, ADF.

Relevance - the construction of models that will help in the study and prediction of processes with long memory.

An analysis of the obtained results was carried out, the analysis of the obtained prediction model was performed.

Ways of further development of the subject of research - development of methods of forecasting and research of processes with a long memory.

PROCESSES WITH LONG MEMORY, ARFIMA, FIAPARCH, HERSTA
INDEX, KPSS, ADF.

ЗМІСТ

ВСТУП	10
РОЗДІЛ 1. ПРОЦЕСИ З ДОВГОЮ ПАМ'ЯТТЮ. ПРИЧИНИ ВИНИКНЕННЯ ТА ПІДХОДИ ДО АНАЛІЗУ	14
1.1. Довга пам'ять	14
1.2. Історія виникнення	15
1.3. Приклади даних	17
1.4 Адекватність моделі	34
Висновки до розділу та постановка задачі	36
РОЗДІЛ 2. МЕТОДИКА МОДЕЛЮВАННЯ ПРОЦЕСІВ ІЗ ДОВГОЮ ПАМ'ЯТТЮ	39
2.1 Особливості структури загальних ймовірнісних моделей	39
2.1.1 Гаусів процес	39
2.1.2 Фрактальний гауссів шум	39
2.1.3 Фрактальний броунівський рух	40
2.1.4 Авторегресійна модель дробно-інтегрованого рухомого середнього	40
2.2. Методи оцінки параметру довгої пам'яті	42
2.3 Методика моделювання процесів із довгою пам'яттю	46
2.3.1 Процес ARIFMA	47
2.3.2 Процеси із довгою пам'яттю в дисперсіях	49
2.3.3 Процеси із подвійною довгою пам'яттю	53
2.3.4 Функції розподілу	55
Висновки до розділу	55
РОЗДІЛ 3. ВИКОНАННЯ ОБЧИСЛЮВАЛЬНИХ ЕКСПЕРИМЕНТІВ	57
3.1 Короткий опис системи R	57
3.2 Приклади побудови моделі	58
3.1 Побудова моделі ARIFMA із потсійною дисперсією	64
3.2 Побудова моделі ARIFMA – FIGARCH/FIAPARCH	67
3.3 Побудова прогнозу дохідності	81
Висновок до розділу	84
РОЗДІЛ 4 РОЗРОБКА СТАРТАП-ПРОЕКТУ	85
4.1 Опис ідеї проекту	86

4.2 Технологічний аудит ідеї проекту	88
4.3 Аналіз ринкових можливостей запуску стартап-проекту.....	89
4.4 Розроблення ринкової стратегії проекту	98
Висновок до розділу 4	102
ВИСНОВОК ПО РОБОТІ.....	103
ПЕРЕЛІК ПОСИЛАНЬ	106
ДОДАТОК А ПРИКЛАД ТЕКСТУ ПРОГРАМИ	108

ВСТУП

Актуальність теми. Бурхливий розвиток фінансових ринків в останні десятиліття супроводжується значним зростанням залучених обсягів капіталу, виникненням великого числа нових фінансових інструментів, збільшенням кількості учасників ринку. Збільшення розміру ризикових позицій компаній і активізація їх діяльності з управління портфелями фінансових активів веде до необхідності розробки нових методів аналізу фінансових ринків, що дозволяють отримувати більш надійні оцінки очікуваних результатів інвестиційної діяльності. Ще 20 років тому емпіричні дослідження в галузі фінансів використовували відносно нескладні аналітичні інструменти. Однак, по мірі стрімкого розвитку комп'ютерних технологій, збільшення доступності довгих рядів даних, розробки сучасних економетричних моделей, призначених безпосередньо для вивчення фінансових ринків, ситуація в цій області почала різко змінюватися. Необхідно підкреслити, що цей прогрес не виявився обмеженим академічними рамками, а навпаки, його результати стали активно впроваджуватися в повсякденну фінансову практику, що в свою чергу стимулювало подальші розробки.

Одним з важливих напрямків досліджень є моделювання динаміки прибутковості і волатильності фондового та валютного ринків. Довгий час було прийнято вважати, що прибутковість фінансових ринків слід процесу "випадкового блукання" і, отже, повністю непередбачувана. Ця точка зору повністю відповідає гіпотезі ефективного ринку, що постулює, що вся інформація про ринок включена в поточну ціну активу. У цьому випадку ціни повинні збігатися зі своїми фундаментальними значеннями, будь-які відхилення від яких пов'язані з процесом надходження новин на ринок, що носять випадковий характер. Однак останнім часом з'являються роботи, в яких ставиться під сумнів адекватність цієї гіпотези реалій фінансового ринку.

Деяка, хоча і незначна (у порівнянні з високими дисперсіями), ступінь передбачуваності прибутковості відзначається багатьма авторами, проте можливість її використання з урахуванням виникаючих трансакційних витрат залишається сумнівною.

Іншим важливим напрямом фінансового моделювання є дослідження динаміки волатильності ринків, яке показало мінливість цієї характеристики в часі, тоді як моделі, що використовувалися раніше, припускали волатильність постійною. Крім інтересу теоретичного, адекватне моделювання волатильності є необхідним також і для практичних додатків в області вимірювання, контролю та управління ризиком. Метою моделювання волатильності є побудова її прогнозу і вивчення різних аспектів ринкової прибутковості. Подібні прогнози застосовуються в таких галузях фінансової діяльності, як ризик менеджмент, оцінка вартості похідних фінансових інструментів, визначення структури портфеля, вибір оптимального часу для здійснення операцій на ринку і т.д.

У кожному з цих випадків важливим виявляється побудова оцінки очікуваної в майбутньому волатильності.

Обидва описаних вище напрямки досліджень продовжують розвиватися, спираючись на новітні розробки в області фінансової економетрики. Прагнення підібрати модель, найбільш повно відповідає реальній поведінці фінансових ринків, і підвищення якості побудованих прогнозів веде до появи як нових класів моделей, так і модифікацій вже існуючих. Одним з найбільш перспективних напрямків роботи по праву вважається побудова моделей з довгою пам'яттю. Ряд з довгою пам'яттю характеризується функцією автокореляції, яка повільно зменшується в міру збільшення тимчасового лага. Протягом багатьох років дослідниками активно використовувалися моделі класу ARMA (процеси з короткою пам'яттю), які передбачають експонентний темп зменшення коефіцієнтів в функції реакції на зовнішні імпульси. В останні півтора десятиліття з'явилося безліч робіт, що підкреслюють важливість ефекту тривалості впливу шоків. Частково інтегровані процеси (процеси з довгою

пам'яттю) можуть розглядатися як проміжний варіант між ставшими вже традиційними альтернативами короткої і нескінченної пам'яті ряду. Таким чином, важливим внеском процесів з довгою пам'яттю в моделювання часових рядів є можливість отримання посвідки функції реакції на імпульси і, отже, довгострокових прогнозів, відмінних від передбачуваних раніше існувавших моделей і більш відповідних емпіричним даним. В останні роки моделі даного типу все більш активно застосовуються до різних рядах фінансових даних, таких як прибутковість фінансових активів, форвардні премії, диференціали процентних ставок, темпи інфляції. Найбільш же успішним на даний момент виявляється використання процесів з довгою пам'яттю для моделювання волатильності цін активів.

Мета дослідження – підвищення якості моделей для процесів із довгою пам'яттю шляхом удосконалення системної методології побудови моделей та комплексного застосування кількох методів інтелектуального аналізу даних. Для досягнення мети потрібно вирішити такі завдання:

а) зібрати статистичні дані із впливом довгої пам'яті для виконання обчислювальних експериментів;

б) вибрати типи математичних моделей і методів інтелектуального аналізу даних для формального опису;

в) побудувати вибрані типи моделей за допомогою статистичних даних та оцінити якість цих моделей;

г) виконати порівняльний аналіз отриманих результатів.

Об'єкт дослідження: процеси з довгою пам'яттю: моделювання і прогнозування.

Предмет дослідження: математичні моделі для опису процесів із довгою пам'яттю, методи їх прогнозування на основі статистичних даних

Методи дослідження: ARFIMA – FIAPARCH, ARFIMA – GARCH, KPSS, ADF.

Наукова новизна одержаних результатів.

1. Удосконалено системну методологію побудови моделей процесів із довгою пам'яттю шляхом корегування структури вибірки даних.
2. Комплексне застосування кількох методів інтелектуального аналізу даних, яке дало можливість підвищити якість прогнозів (загальну точність).

Пояснювальна записка складається з чотирьох розділів. У першому розділі висвітлено актуальність дослідження процесів із довгою пам'яттю, ознайомлення з теорією, означено основні поняття. У другому розділі розглянуто ряд математичних моделей, за допомогою яких будують прогнози. Третій розділ описує процес побудови моделей, також у ньому аналізуються результати роботи цих моделей. Четвертий розділ присвячений розробці стартап-проекту за матеріалами дослідження.

РОЗДІЛ 1. ПРОЦЕСИ З ДОВГОЮ ПАМ'ЯТТЮ. ПРИЧИНИ ВИНИКНЕННЯ ТА ПІДХОДИ ДО АНАЛІЗУ

1.1. Довга пам'ять

Довга пам'ять, або довгострокова залежність - властивість, що описує кореляційну структуру високого порядку тимчасового ряду. У разі, якщо ряд володіє довгою пам'яттю, то залежність існує навіть між далеко розташованими один від одного в часі спостереженнями. Оскільки довга пам'ять створює в моделі для середнього рівня ряду нелінійну залежність в перших моментах розподілу і, отже, генерує в динаміці ряду потенційно придатний для прогнозування компонент, її присутність в рядах прибутковості фінансових активів породжує сумніви в ефективності фінансового ринку. Ціна активу на ефективному ринку повинна слідувати мартінгальному процесу, в якому кожне цінове зміна не піддається впливу своїх попередників, тобто не володіє пам'яттю.

Оскільки довга пам'ять має на увазі наявність значущої автокорреляції між віддаленими спостереженнями, її присутність веде до того, що минулі прибутковості можуть допомогти в передбаченні доходностей майбутніх, що в свою чергу призводить до можливості отримання спекулятивного прибутку. Наявність такої структури в дохідні активи піднімає ряд теоретичних і емпіричних питань. У зв'язку з тим, що довга пам'ять являє собою особливу форму нелінійної динаміки, вона ставить під сумнів лінійне моделювання і вимагає розробки на теоретичному рівні нових нелінійних моделей оцінки фінансових активів для обліку поведінки, пов'язаного з наявністю довгою пам'яті. Мандельброт (1971) зазначив, що в присутності довгої пам'яті вплив надходження нової інформації на ринок не може бути повністю усунуто внаслідок арбітражу, і мартінгальні моделі ціноутворення не можуть бути виведені з арбітражних міркувань. Більш того, ціноутворення похідних цінних паперів мартінгальними методами не буде коректним у разі, якщо покладений в його основі стохастичний процес має довгу пам'яттю. Статистичні висновки, що

стосуються моделі ціноутворення активів, що базуються на стандартних тестових процедурах, також можуть бути незастосовні для рядів з довгою пам'яттю.

1.2. Історія виникнення

Задовго до того, як були доступні відповідні стохастичні процеси, спостерігалися відхилення від незалежності, що були помітні далеко за межі звичайного часового горизонту, часто навіть у ситуаціях, коли незалежність здавалася б природним припущенням.

Наприклад, канадсько-американський астроном і математик Саймон Ньюкомб (Newcomb 1895) помітив, що в астрономії помилки зазвичай впливають на цілі групи послідовних спостережень і різко збільшують "ймовірну помилку" оцінюваних астрономічних констант, так що звичайне σ/\sqrt{n} - правило більше не застосовується.

Хоча існує цілий ряд можливих причин для якісного пошуку Ньюкомба, стаціонарні процеси довготривалої пам'яті надають правдоподібне "пояснення". Подібні висновки були зроблені раніше Пірсом (1873) (див. Також обговорення даних Пірса Вілсона та Хільфerti 1929 р., А пізніше - в книзі Мостеллера і Тукі 1977 р. У розділі під назвою «Як σ / \sqrt{n} може ввести в оману»). Зауваження Ньюкомба були підтверджені кілька років пізніше Карлом Пірсоном (1902), який проводив експерименти, що імітують астрономічні спостереження. Використовуючи складну експериментальну установку, він продемонстрував не тільки те, що спостерігачі мали власне упередження, але також кожна окрема серія вимірів показала постійні послідовні кореляції.

Для просторових даних австралійський агроном Х. Фейрфілд Сміт (1938) винайшов у випробуваннях рівномірності емпіричний закон для виміру

врожайності пшениці, що суперечить припущенню про незалежність або сумованність кореляції, оскільки стандартне відхилення середнього зразка збігається з нулем з більш повільною швидкістю, ніж корінь квадратний з розміру земельної ділянки.

Ці висновки були згодом підхоплені Уїтлом (1956, 1962), який запропонував модель простору-часу на основі стохастичних диференціальних рівнянь, що демонструють гіперболічні розпадаються просторові кореляції і тим самим можливого пояснення емпіричного закону Фейрфілд Сміта.

У гідрології Херст (1951) відкрив емпіричний закон при вивченні довготривалого зберігання резервуарів для Нілу. Опираючись на його емпіричні висновки, Херст рекомендував збільшити висоту запланованої Асуанської високої греблі далеко за рамки звичайних прогнозів. Феллер (1951) показав, що висновки Херста несумісні з припущеннями про слабку залежність чи кінцеві моменти. Пізніше Мандельброт створив терміни "ефект Ноя" для довгохвильових розподілів та ефект Джозефа або Херста для "довготривалої залежності". Останнє стосується Буття 41, 29-30, де "сім років великого достатку" і "сім років голоду" можна інтерпретувати як свідчення сильних послідовних кореляцій. Підхід Мандельброта і його співробітників веде до нової галузі математики, яка замінила звичайні геометричні об'єкти, «фрактал» і «самоподібності» (наприклад, Мандельброт 1965, 1967 1969 1971, 1977, 1983; Мандельброт і Ван-Несс 1968, Мандельброт і Уолліс 1968a, b, 1969a, b, c), і популяризацію теми у багатьох наукових областях, включаючи статистику. У економіці явище довгої пам'яті було відкрито Грейнджером (1966). Одночасно з Хоскінгом (1981), Грейнджер (1980) представив дрібні моделі ARIMA, які значно покращили придатність залежності довгого діапазону в статистичній практиці.

У геології Мацєрон розробив поле геостатистики, використовуючи, зокрема, процеси та статистичні методи моделювання просторової довгої пам'яті (див., Наприклад, Matheron 1962, 1973, Solo 1992). З математичної точки зору,

основні поняття фракталів, самоподібності та довготривалої залежності існували задовго до того, як тема стала модною, однак їх практичне значення не було повністю визнано до піонерської роботи Мандельброта. Наприклад, розмірність Хаусдорфа, який грає ключову роль у визначенні фракталів, була введена Хаусдорфом (1918) і детально вивчена Безіковичем (наприклад Безіковичем 1929 року, Безікович і Ерселлом 1937). У 17ст. Лейбніц (1646-1716) вивчав рекурсивну самоподібність, і приблизно сто років по тому Карл Вейерштрасс описав функцію, яка є безперервною, але ніде не відрізняється. Перший фрактал віднесено чеському математику Бернарду Больцано (1781-1848). Інші ранні фрактали включають множину Кантора (1883), криву Коха (Хельге фон Кох 1904), трикутник Серпінського (1915) та криву Леві (1938).

1.3. Приклади даних

У цьому розділі наведені приклади процесів із довгою пам'яттю, а також буде введено кілька евристичних методів для виявлення та оцінки сили довготривалої залежності. Найбільш поширені галузі, де довготривала залежність відбувається найчастіше - дендрохронологія та гідрологія. Тому ми почнемо з прикладів із цих областей. Щорічні вимірювання кількості кілець дерева зазвичай простягаються протягом сотень років, і довга пам'ять часто зустрічається в досить "чистій" формі в тому сенсі, що гіперболічна поведінка автокореляцій і спектральної щільності виконується практично для всіх лагів та частот відповідно.

Тому серії кілець дерев часто використовуються як основні приклади сильної залежності та самоподібності. Розглянемо, наприклад, Рисунок 1.1 (джерелом даних є Hyndman, бібліотека даних часових рядів, <http://robjhyndman.com/TSDL>). Можна спостерігати наступні характерні риси:

1) **Помилкові тенденції, цикли та самоподібність.** Спостережувана серія демонструє локальні тенденції та періодичності, які здаються хибними, оскільки вони зникають та мають різну тривалість і частоту. Крім того, ці особливості та загальне візуальне враження від часових рядів залишаються незмінними при розгляді агрегованих даних з непересічними сусідніми блоками (див. Рисунок 1.2). Це вказує на стохастичну "самоподібність", яка є властивістю того, що масштабування часу змінює розподіл імовірності лише масштабним коефіцієнтом.

Примітка: Масштабний коефіцієнт – це число, яке масштабує або помножує деяку кількість. У рівнянні $y = Cx$, C – коефіцієнт масштабу для x .

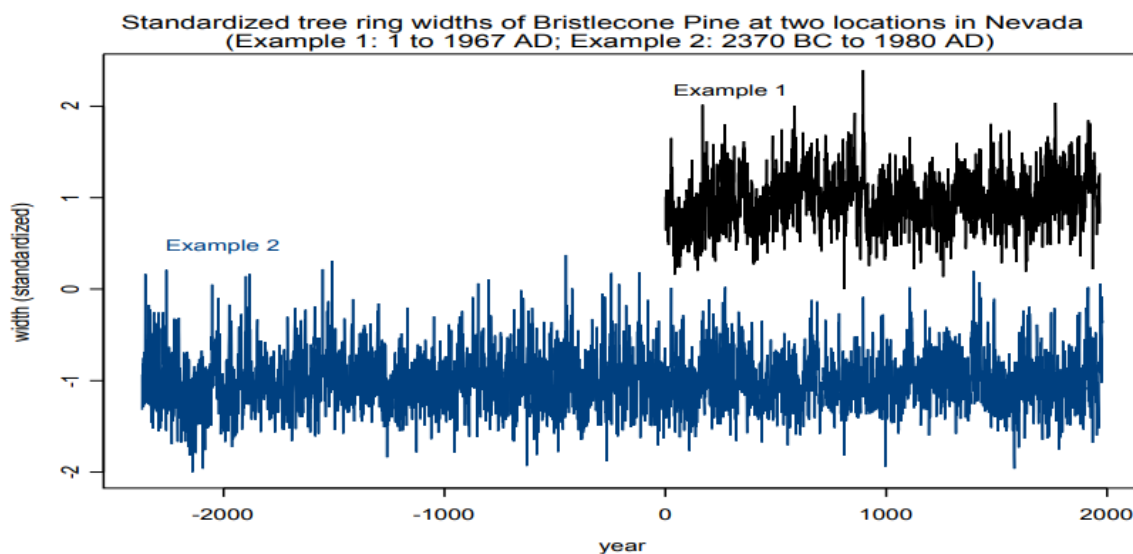


Рисунок 1.1 – Дві типові серії кілець дерева

2) **Повільне гіперболічне згасання.** Зразки автокореляцій повільно згасають із зростанням затримки k .

$$\hat{p}(k) = \frac{1}{n} \sum_{i=1}^{n-|k|} (x_i - \bar{x}) (x_{i+|k|} - \bar{x})$$

Більш детально – згасання $\hat{p}(k)$ виявляється гіперболічним зі швидкістю k – α (для деяких $0 < \alpha < 1$), мається на увазі, що його неможливо просумувати. В різних джерелах це явище називається довгою пам'яттю, сильною пам'яттю, пам'яттю далекої залежності або довготривалими кореляціями. Це ілюструється на Рисунку 1.3-с), де $\log \hat{p}(k)$ протиставляється $\log k$. Точки знаходяться безпосередньо навколо $\log \hat{p}(k) \approx \text{const} + \beta_p \log k$, де $\beta_p \approx -0,5$. Так само дисперсія вибіркового середнього згасає до 0 повільніше ніж n^{-1} . Це можна дослідити емпірично і побачити на Рисунку 1.3-d) $\log s_m^2$ vs. $\log m$, де s_m^2 – це виборочна дисперсія на основі непересічних блоків m спостережень, тобто:

$$s_m^2 = \frac{1}{n_m - 1} \sum_{i=1}^{n_m} (\overline{x_{(i-1)m,m}} - \bar{x})^2$$

де

$$\overline{x_{t,m}} = \frac{1}{m} \sum_{j=1}^m x_{t+j}$$

і $n_m = [n / m]$. Результируючий «схил» на Рисунку 1.3-d) наближається до $\beta_{s^2} = -0,4$, роблячи s_m^2 пропорційним до $m^{-0,4}$, що набагато повільніше за звичайний розмір m^{-1} .

Наступний метод, який іноді використовується для виявлення довгострокових залежностей, є так званий R/S – аналіз, зображення якої можна спостерігати на Рисунку 1.3-f). R/S-аналіз — це сукупність статистичних прийомів та методів аналізу часових рядів (здебільшого фінансових), що дозволяють визначити деякі важливі їх характеристики, такі як наявність неперіодичних циклів, пам'яті та інших.

Історія створення методології R/S – аналізу бере початок з середини XX-го століття, коли гідролог Херст, проробивши майже 40 років над проектом

Нільської греблі, завершував обробку часових рядів об'ємів стоку річок. Коли Херст вирішив перевірити припущення про те, що ці ряди підкоряються нормальному закону, він в результаті винайшов нову статистику - показник Херста (H). Як виявилось, цей показник має широке застосування в аналізі часових рядів завдяки своїй чудовій стійкості. Він містить мінімальні припущення про систему, що вивчається, і може класифікувати часові ряди. Він може відрізнити випадковий ряд від не випадкового, навіть якщо випадковий ряд не гаусівський (тобто не нормально розподілений). Херст виявив, що більшість природних систем не характеризуються випадковим блуканням - гаусівським, тобто поведінка часових рядів показників цих систем не підкоряється нормальному закону. Цей факт означає непридатність інструментарію економетрики для статистичного аналізу природних часових рядів. Херст вимірював коливання води в резервуарі відносно середнього з плином часу і ввів безрозмірне відношення за допомогою ділення розмаху R на стандартне відхилення спостережень S. Цей спосіб аналізу почав називатися методом нормованого розмаху (R/S-аналізу). Херст показав, що більшість природних явищ, включаючи річкові стоки, температуру, осідання, сонячні плями характеризуються "зміщеним випадковим блуканням" – трендом із збуренням. Сила тренду та рівень збурення можуть бути оцінені тим, як змінюється нормований розмах з часом, або, іншими словами, на скільки величина H перевершує 0,5. Приведемо опис алгоритму R/S – аналізу в тому вигляді, як ми будемо його реалізовувати.

$$R/S(t, m) = \frac{R(t, m)}{S(t, m)};$$

де

$$R(t, m) = \max_{1 \leq i \leq m} \left(y_{t+i} - y_t - \frac{i}{m} (y_{t+m} - y_t) \right) - \min_{1 \leq i \leq m} \left(y_{t+i} - y_t - \frac{i}{m} (y_{t+m} - y_t) \right),$$

$$y_u = \sum_{i=1}^u x_i,$$

i

$$S(t, k) = \sqrt{\frac{1}{m} \sum_{i=t+1}^{t+m} (x_i - \bar{x}_{t,m})^2}.$$

Це визначення походить з гідрології, де R відповідає оптимальній ємності резервуара, коли відтік лінійний, а x_i відображає надходження води в час i . Рисунок 1.3-f) показує та порівнює $R / S(t, m)$ проти m в логарифмічному масштабі по обом осям. Знову ж таки, ми бачимо лінійний зв'язок між $\log R / S$ (як функції від m), та $\log m$, з нахилом близьким до $\beta_{R/S} = 0,8$, на відміну від залежності, або короткотривалої залежності, де очікуватиметься нахил 0,5.

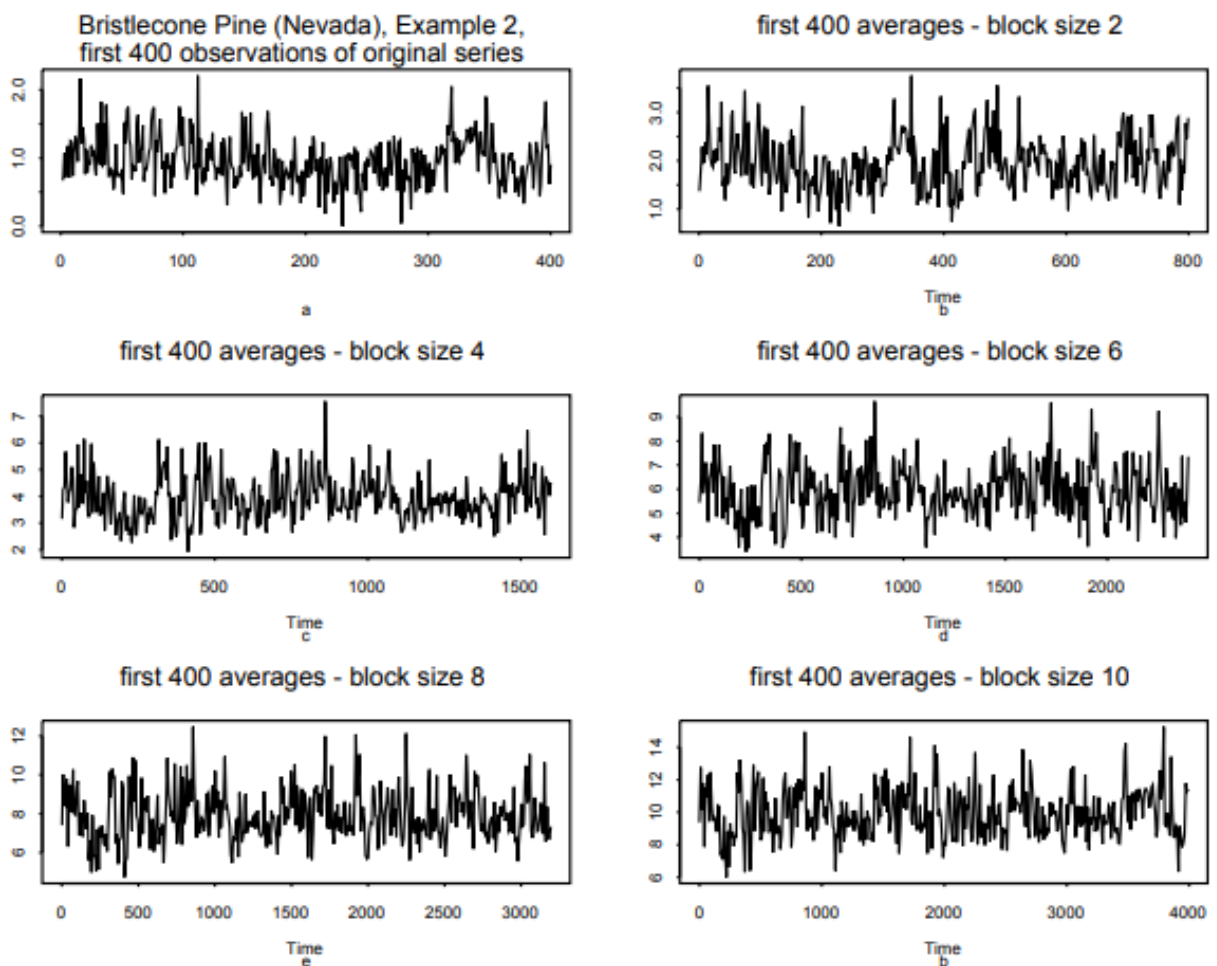
І, нарешті, Рисунок 1.3-f) зображує логарифм періодограми $I(\lambda)$ (як емпірична аналогія спектральної щільності f) та протиставляє його частоті вимірювань. Негативний нахил близьк 0,5, свідчить що полюс спектральної щільності знаходиться у положенні порядку $\lambda^{-0,5}$.

Схожі результати спостерігалися у Прикладі №2, на малюнках від 1.4-a) до 1.4-f).

Далі ми розглянемо та проаналізуємо дані річкових течій. Рисунки 1.5-a), 1.6-a), 1.7-a) та 1.8-a) показують середньомісячну річковий ріст (в м 3 / с) для чотирьох річок з різних частин світу: 1) Маас на станції Lith (Нідерланди) ; 2) Вісла на Тчеве (Польща); 3) Тежу у В.В. де Родао (Португалія) і 4) Біла Річка в гирлі біля Оурі, штат Юта (США). Дані наведено в базі даних про ріні річок Центру стійкості та глобального навколишнього середовища, Інституту екологічних досліджень Гейлорда Нельсена, Університет Вісконсін-Медісон. Оскільки це щомісячні дані,

існує сильний сезонний компонент. Щоб отримати уявлення про структуру залежності, сезонний ефект спочатку видаляється шляхом віднімання відповідних щомісячних значень (тобто середньої температури січня, середньої температури лютого тощо). Оригінальні та семантичні дані відображаються у верхній та нижній частині кожної фотографії тимчасової серії відповідно. Для кожної з сезонно-орієнтованих серій точки в журнальній журналі періодограми (всі блоки b) розміщені навколо прямої лінії для всіх частот.

Рисунок 1.2 – а) Ряд кілець дерева, Приклад 1; b) - f) – сукупна серія $\bar{x}_t =$



$m^{-1}(x_{(t-1)m+1} + \dots + x_{tm}), t = 1, 2, \dots, 400$ з довжиною блоків 2, 4, 6, 8 і 10 відповідно.

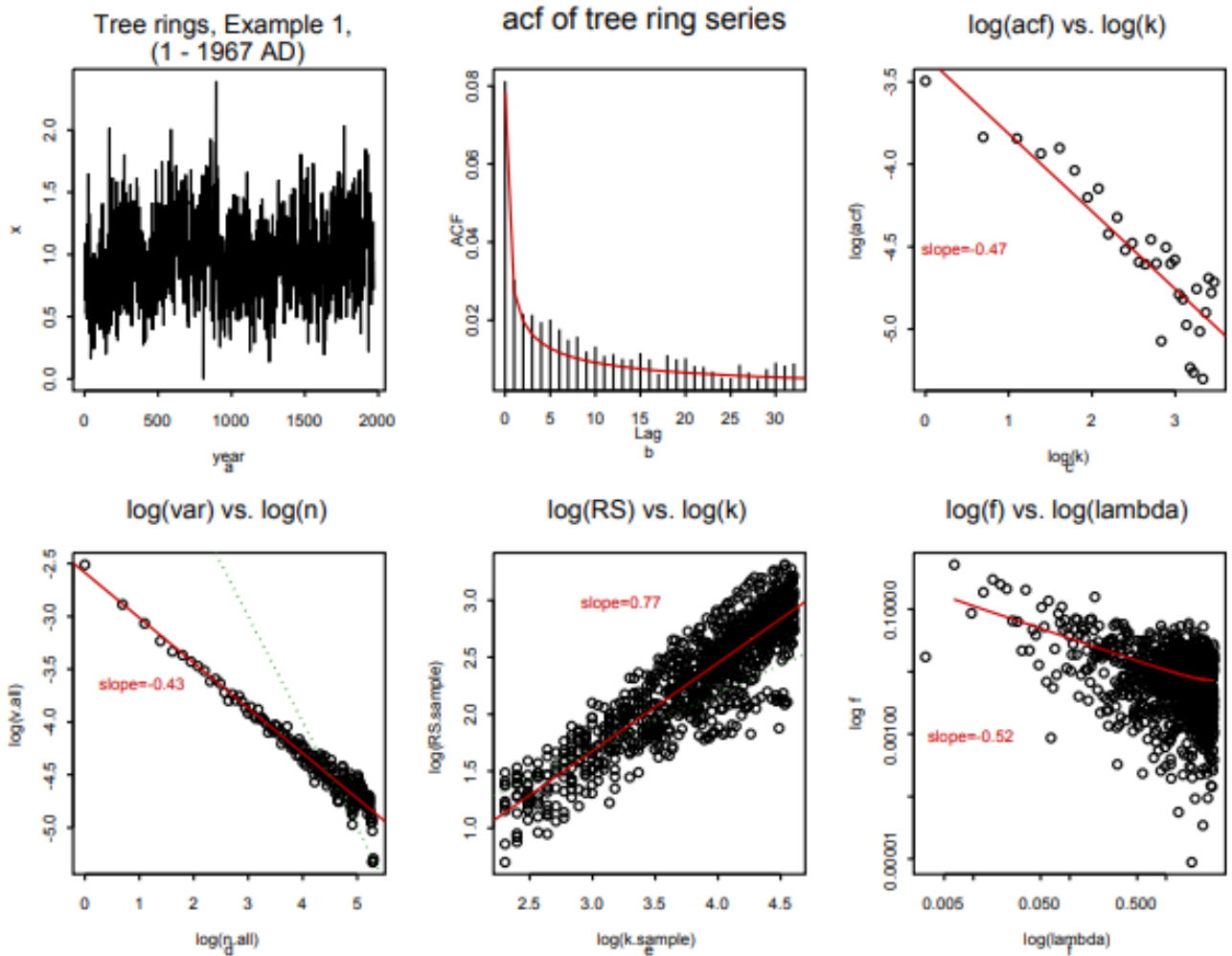


Рисунок 1.3 – Дослідження кілець дерева №1: а) щорічно спостережувана серія; б) емпіричні автокореляції $\hat{p}(k)$; в) $\log \hat{p}(k)$ vs. $\log k$; г) $\log s_m^2$ vs. $\log m$; д) $\log R/S$ vs. $\log k$; е) $\log I(\lambda)$ vs. $\log \lambda$

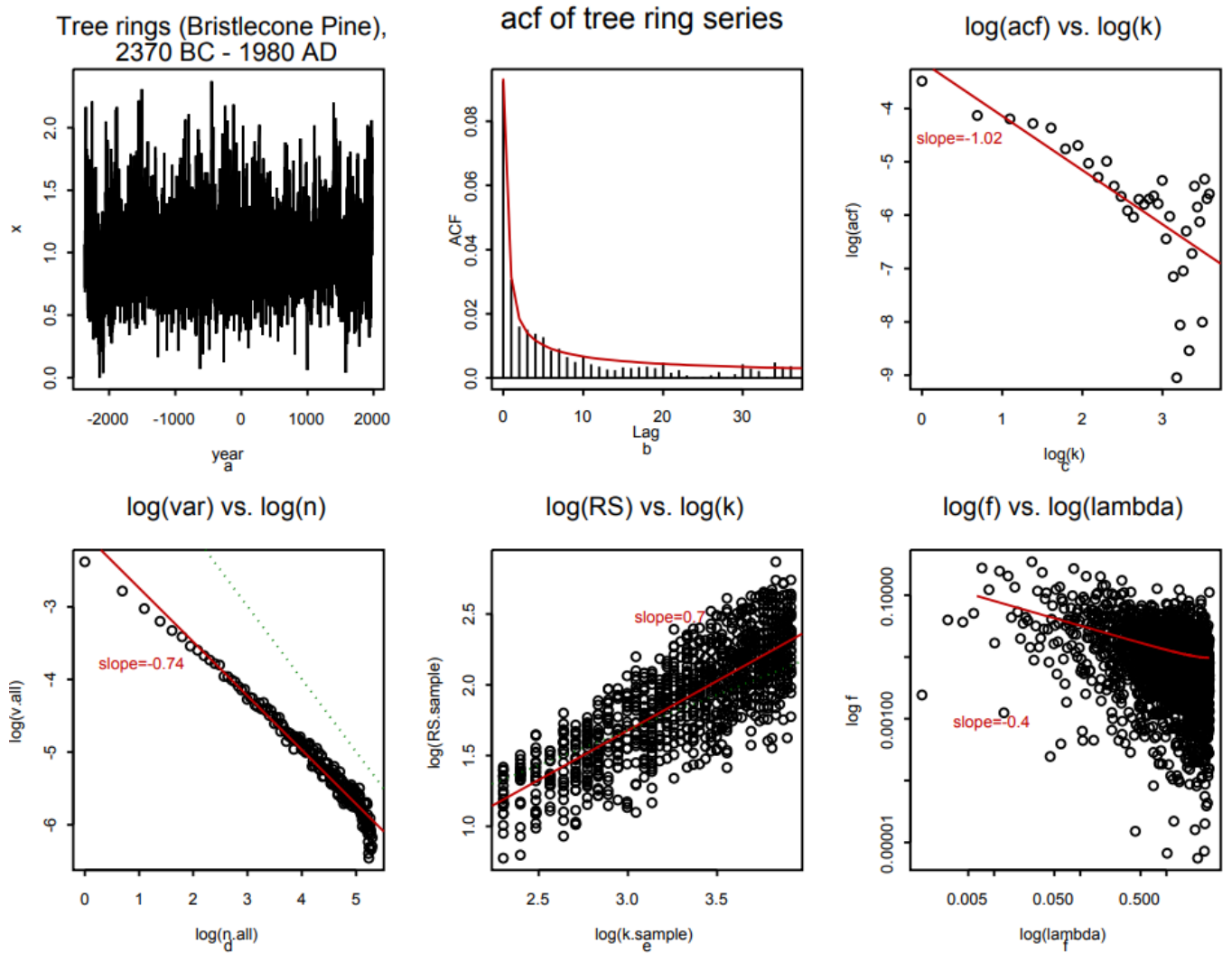


Рисунок 1.4 – Дослідження кілець дерева №2: а) щорічно спостережувана серія; б) емпіричні автокореляції $\hat{p}(k)$; в) $\log \hat{p}(k)$ vs. $\log k$; д) $\log s_m^2$ vs. $\log m$; е) $\log R/S$ vs. $\log k$; ф) $\log I(\lambda)$ vs. $\log \lambda$

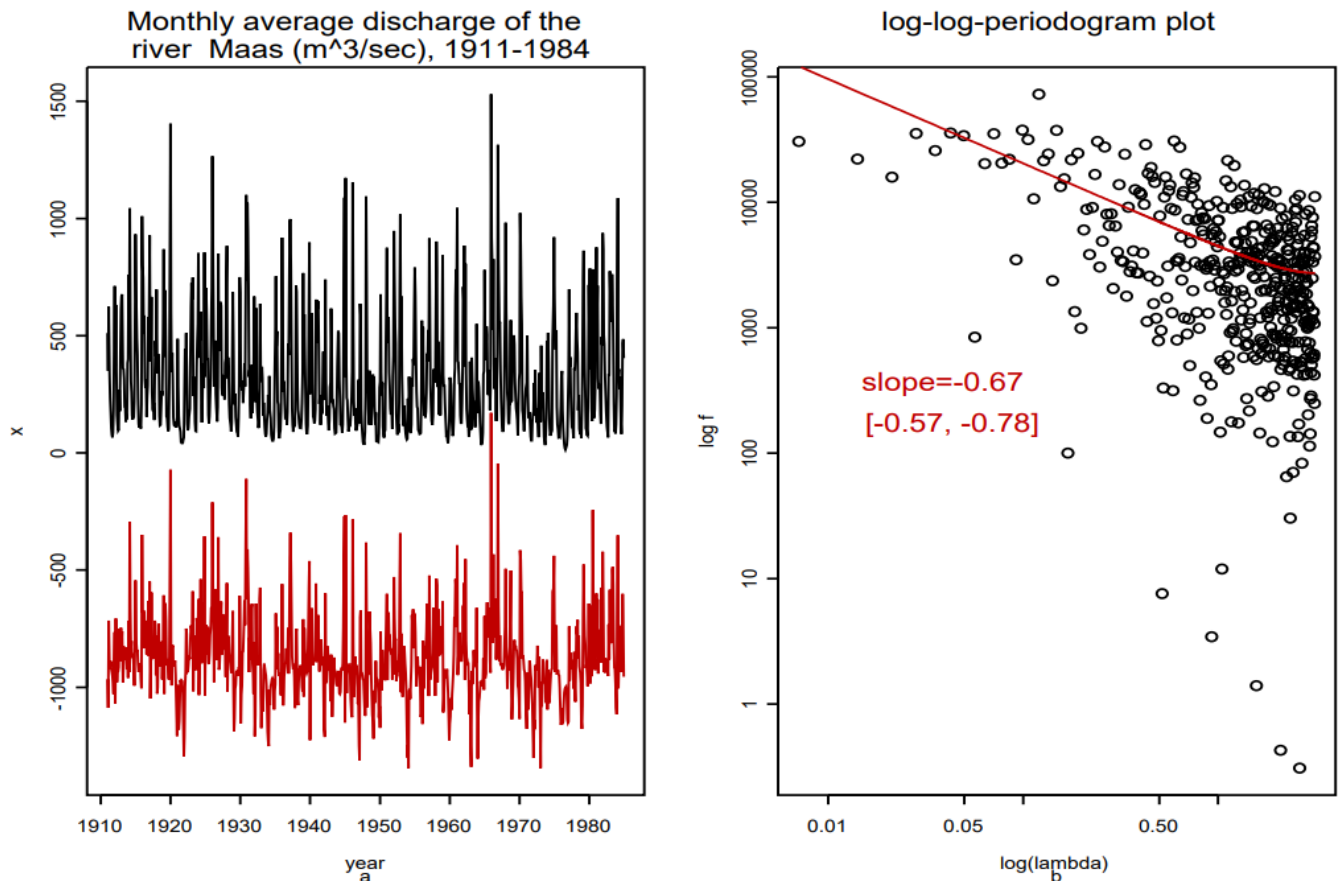


Рисунок 1.5 – а) Середня щомісячна інформація стоку річки Маас, б) «десезонована» періодограма вимірювань

Наведені приклади даних можуть бути дещо оманливими, оскільки може створитися враження, що відкриття тривалої пам'яті можна легко виконати шляхом встановлення прямої лінії до результатів спостережень у відповідному логарифмічному графіку. На жаль, ситуація ускладнюється, навіть якщо врахувати лише річкові стоки. Наприклад, на рисунках 1.9, 1.10 та 1.11 показані логарифмічні графіки для Дунаю на чотирьох різних станціях: 1) Братислава (Словаччина); 2) Надямарос (Угорщина); 3) Дробета-Турну Северин (Румунія); 4) Чітала Ізмаїл (Румунія).

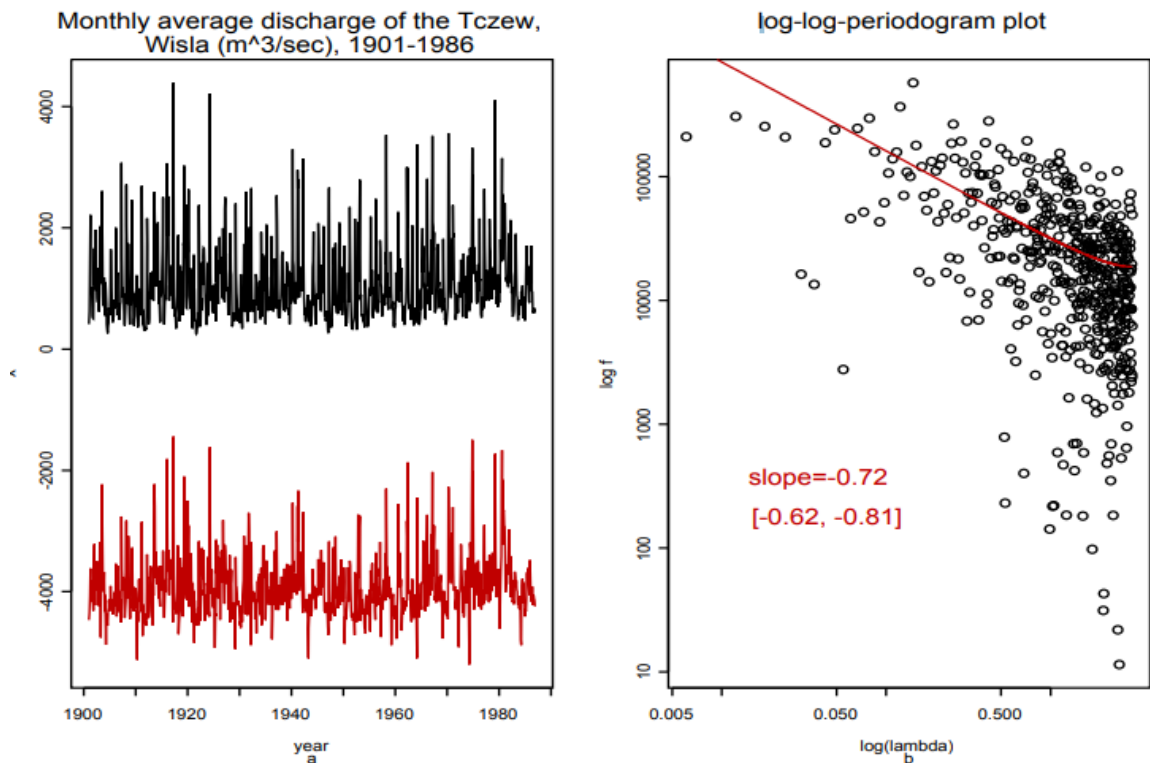


Рисунок 1.6 – а) Середня щомісячна інформація стоку річки Вісла, б) «десезонована» періодограма вимірювань

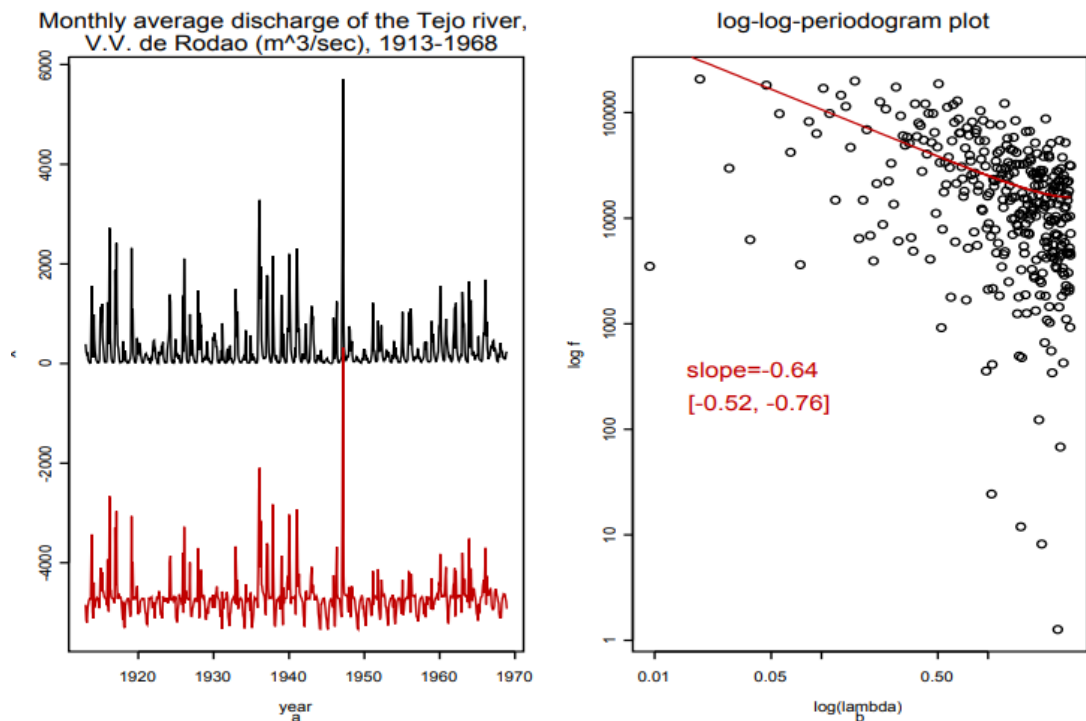


Рисунок 1.7 – а) Середня щомісячна інформація стоку річки Тейо, б) «десезонована» періодограма вимірювань

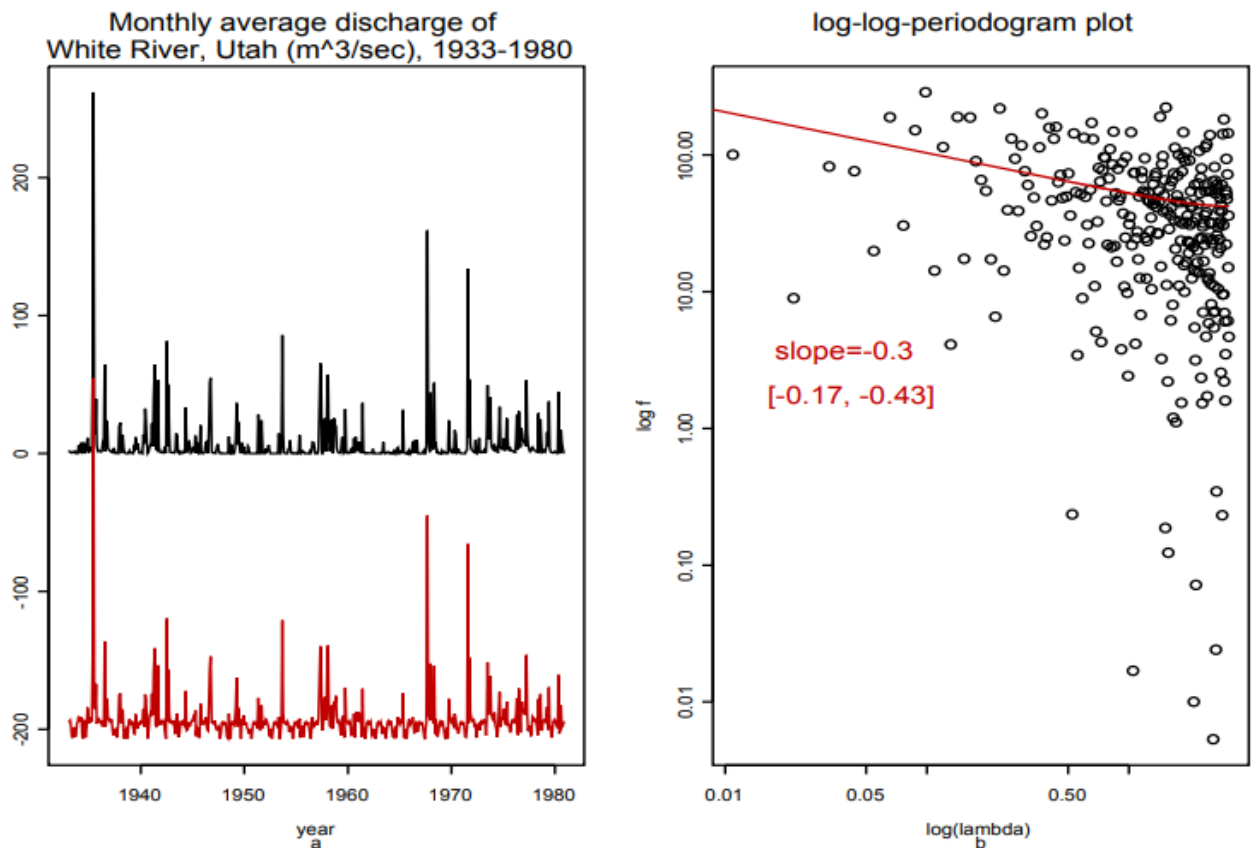


Рисунок 1.8 – а) Середня щомісячна інформація стоку Білої Річки, б) «десезонована» періодограма вимірювань

Розглянемо спочатку вимірювання в Братиславі. Точки логарифмічного графіку більше не збігаються до певної прямої на всьому шляху вимірювань, тому незрозуміло, як оцінити - «ультраамтивні» схили (тобто асимптотичні схили для $m, k \rightarrow \infty$ і $\lambda \rightarrow 0$). Знаходження прямої для всіх точок призводить до низького рівня відповідності (тобто для k та m великі, а λ малі). Це одна з основних проблем при роботі з довгою пам'яттю. Визначення "довгої пам'яті" асимптотичне, тому часто важко виявити та кількісно визначити кінцеві зразки, і значна частина літератури, що описує процеси з довгою пам'яттю стосується цього питання. На відміну від прямих на Рисунках 1.9-b) та 1.9-c), на 1.9-d) спектральна щільність базується на більш складному методі, який поєднує в собі метод максимальної правдоподібності та Баєсів інформаційний критерій (БІК) для фрактальних ARIMA моделей.

Метод максимальної правдоподібності (також метод найбільшої вірогідності) у математичній статистиці — це метод оцінювання невідомого

параметра шляхом максимізації функції правдоподібності. Він ґрунтується на припущенні про те, що вся інформація про статистичну вибірку міститься у цій функції. Метод максимальної правдоподібності був проаналізований, рекомендований і значно популяризований Р. Фішером між 1912 і 1922 роками (хоча раніше він використовувався Гаусом, Лапласом і іншими). Оцінка максимальної правдоподібності є популярним статистичним методом, який використовується для створення статистичної моделі на основі даних, і забезпечення оцінки параметрів моделі. Метод максимальної правдоподібності відповідає багатьом відомим методам оцінки в області статистики. Наприклад, припустимо, що ви зацікавлені зростом мешканців України. Припустимо, у вас дані стосовно зросту деякої кількості людей, а не всього населення. Крім того передбачається, що зріст є нормально розподіленою величиною з невідомою дисперсією і середнім значенням. Вибіркові середнє значення і дисперсія зросту є максимально правдоподібними до середнього значення і дисперсії всього населення.

Для фіксованого набору даних і базової імовірнісної моделі, використовуючи метод максимальної правдоподібності, ми набудемо значень параметрів моделі, які роблять дані «ближчими» до реальних. Оцінка максимальної правдоподібності дає унікальний і простий спосіб визначити рішення у разі нормального розподілу.

БІК — статистичний критерій для обирання моделі серед скінченної множини моделей; найприйнятнішою є модель із найнижчим БІК. Він ґрунтується, зокрема, на функції правдоподібності, і тісно пов'язаний з інформаційним критерієм Акаїке (ІКА).

При пристосовуванні моделей можливо підвищувати правдоподібність шляхом додавання параметрів, але це може призводити до перенавчання. Як БІК, так і ІКА намагаються розв'язувати цю проблему введенням члена штрафу для числа параметрів у моделі; член штрафу в БІК є більшим, ніж в ІКА.

БІК було розроблено Гідеоном Шварцем, і опубліковано в праці 1978 року, в якій він навів баєсівське обґрунтування його застосування.

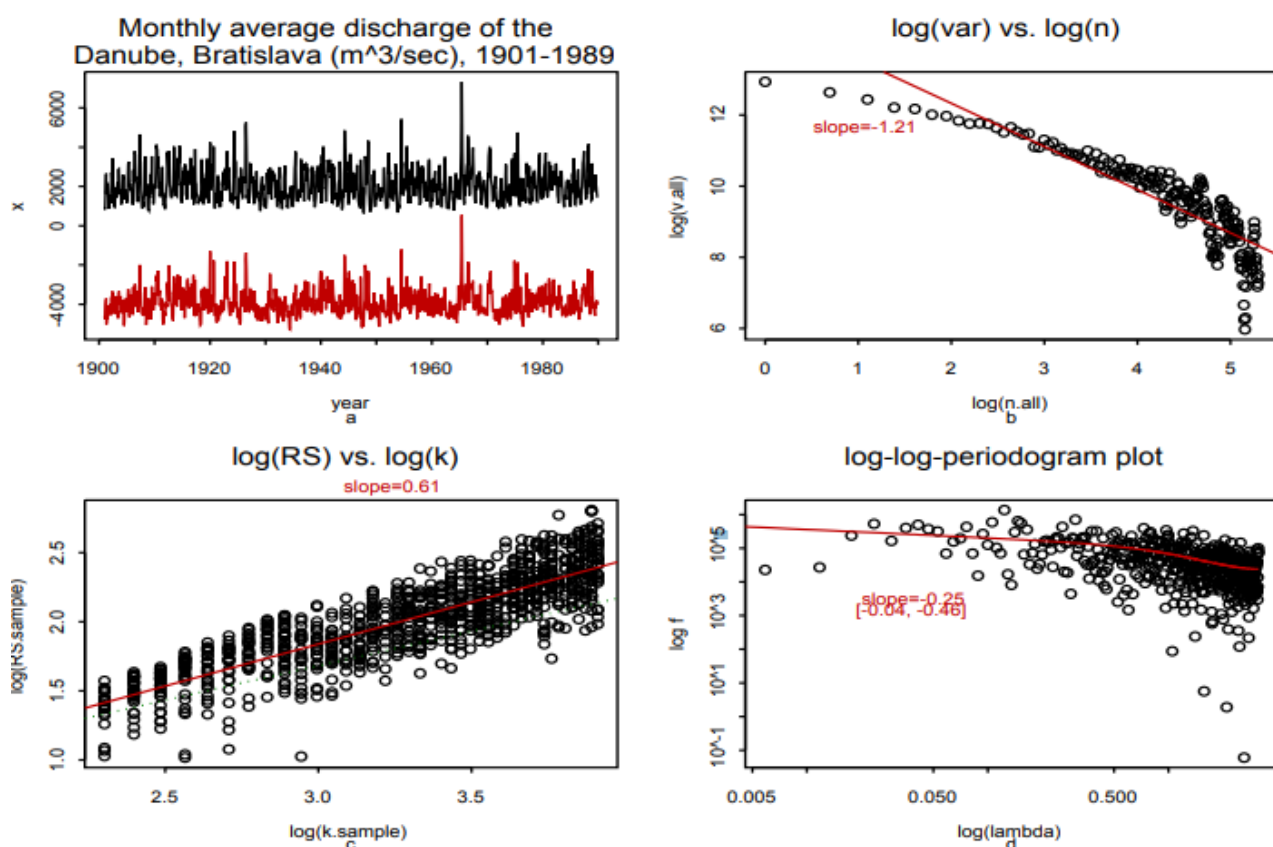
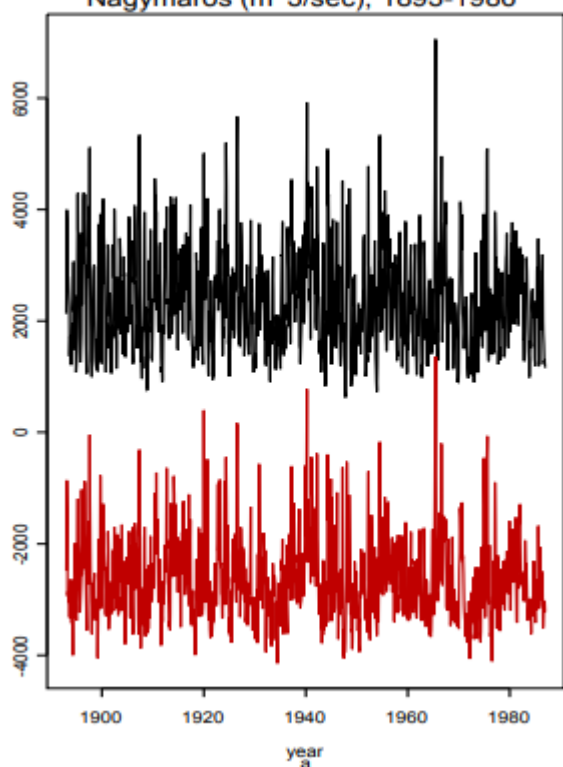


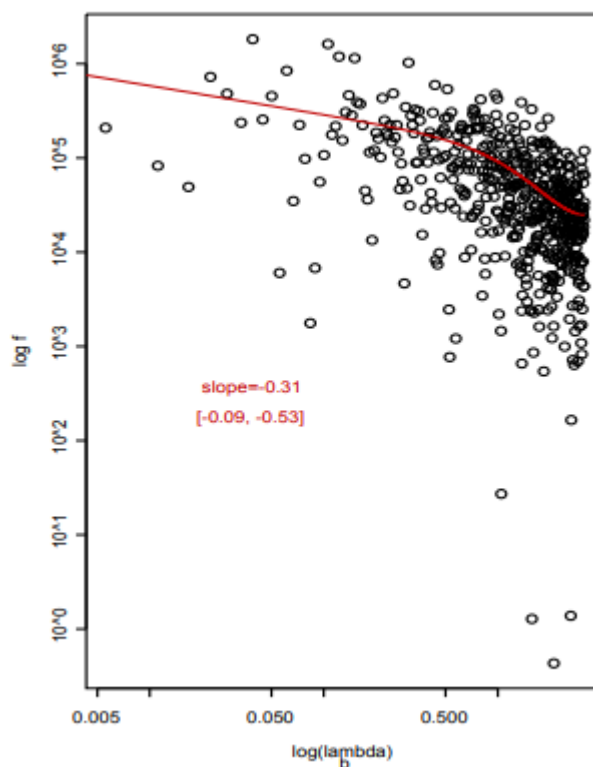
Рисунок 1.9 – Вимірювання стоку Дунаю (Братислава)

Аналогічні спостереження можна зробити і на прикладах інших серій вимірювання рівня Дунаю (Рисунок 1.10, 1.11). Тут також треба звернути увагу на те, що метод максимальної максимальної правдоподібності (MLE) оцінює практично однакові значення для β_f ($-0,25$; $-0,31$; $-0,25$; $-0,29$), це свідчить що значення близько $-0,25$ або $-0,3$ можна вважати типовими для Дунаю у цих регіонах. З іншого боку, кут схилу змінюється в залежності від переміщення місця дослідження вгору або вниз за течією. Наприклад, у Хофкірхені, Німеччина (Рисунок 1.11), явище довгої пам'яті виглядає набагато сильнішим із коефіцієнтом β_f приблизно $-0,75$, а відповідну пряму ми можемо спостерігати на всіх частинах графіку.

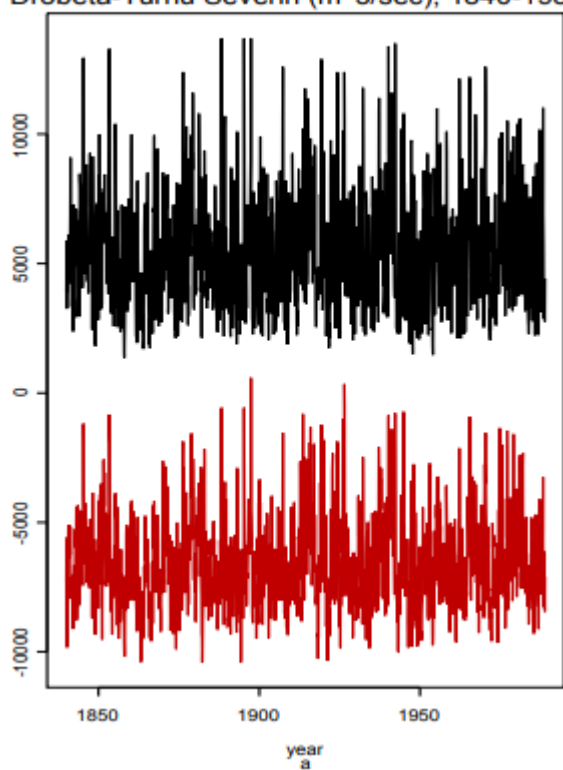
Monthly average discharge of the Danube,
Nagymaros (m³/sec), 1893-1986



log-log-periodogram plot



Monthly average discharge of the Danube,
Drobeta-Turnu Severin (m³/sec), 1840-1988



log-log-periodogram plot

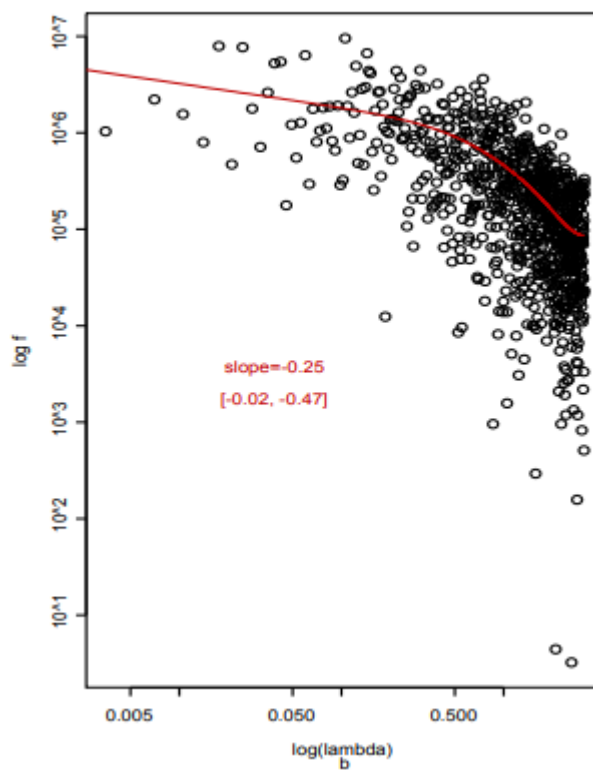


Рисунок 1.10 – Вимірювання стоку Дунаю

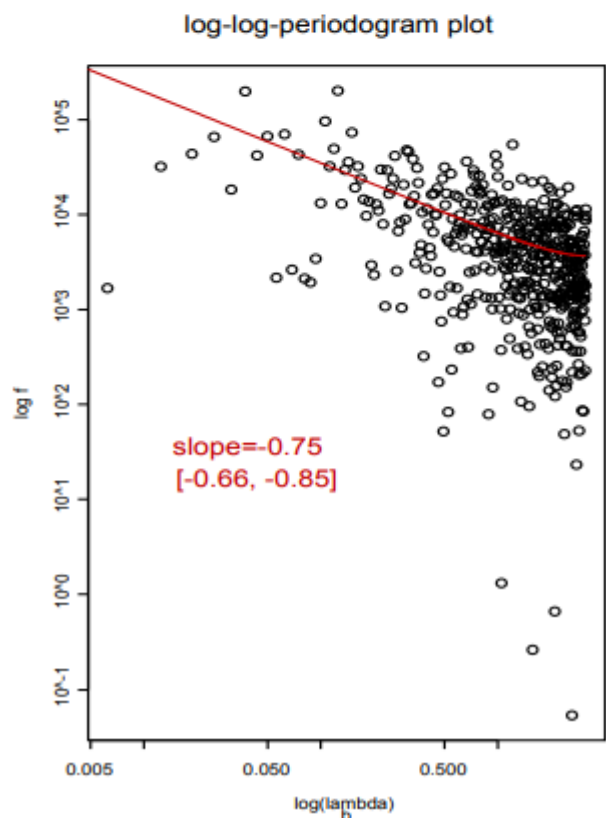
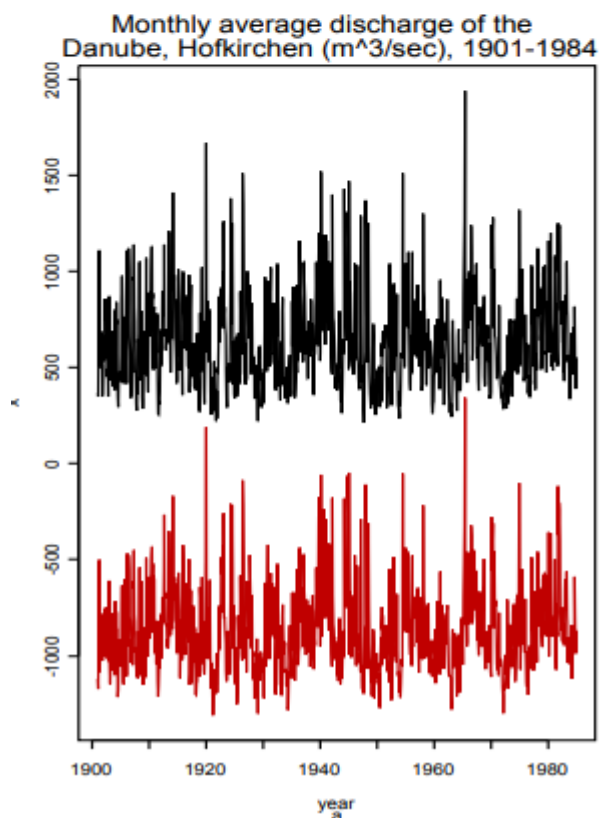
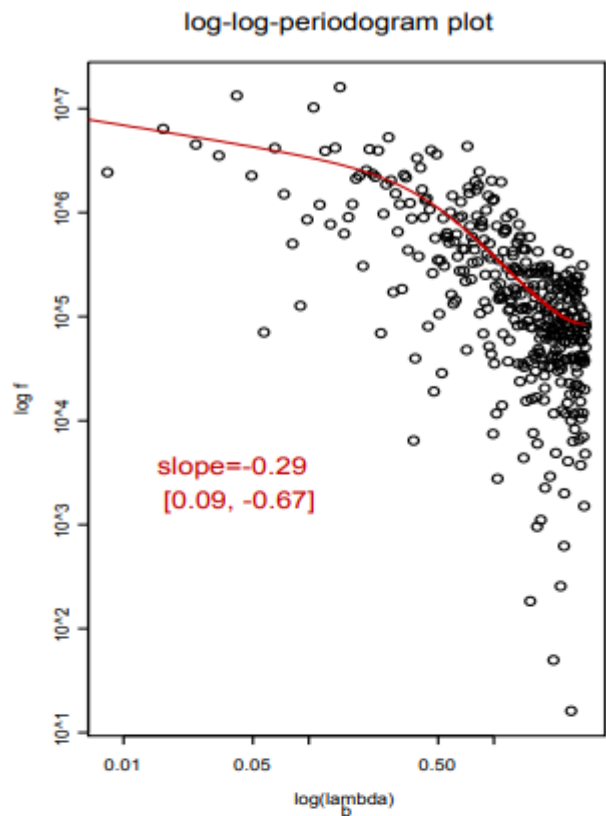
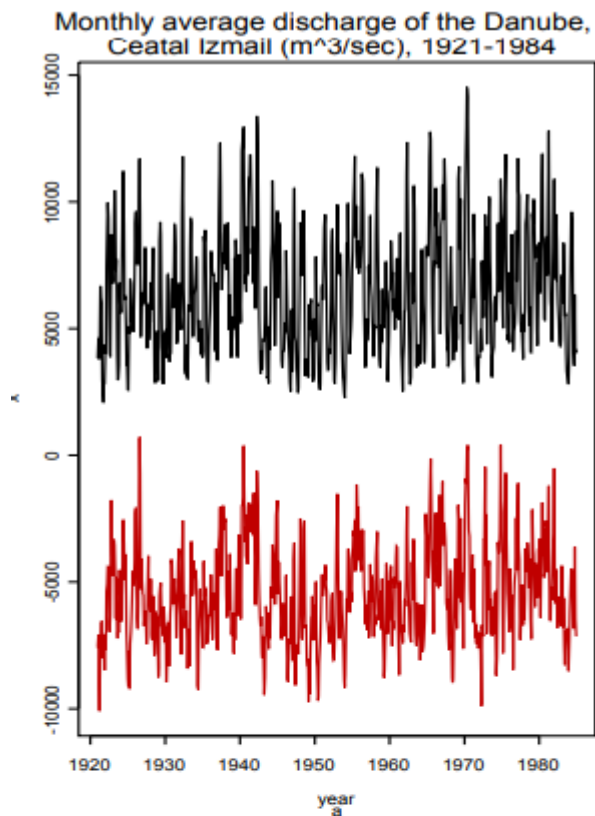


Рисунок 1.11 – Вимірювання стоку Дунаю

Ще більш заплутану серію вимірювань стоку річки Ніл можна спостерігати на графіках Рисунку 1.12. Сезонність як явище в цьому випадку має ще більший вплив на вимірювання, і аналогічні способи уникання цього впливу, як в попередніх прикладах, не допомагає.

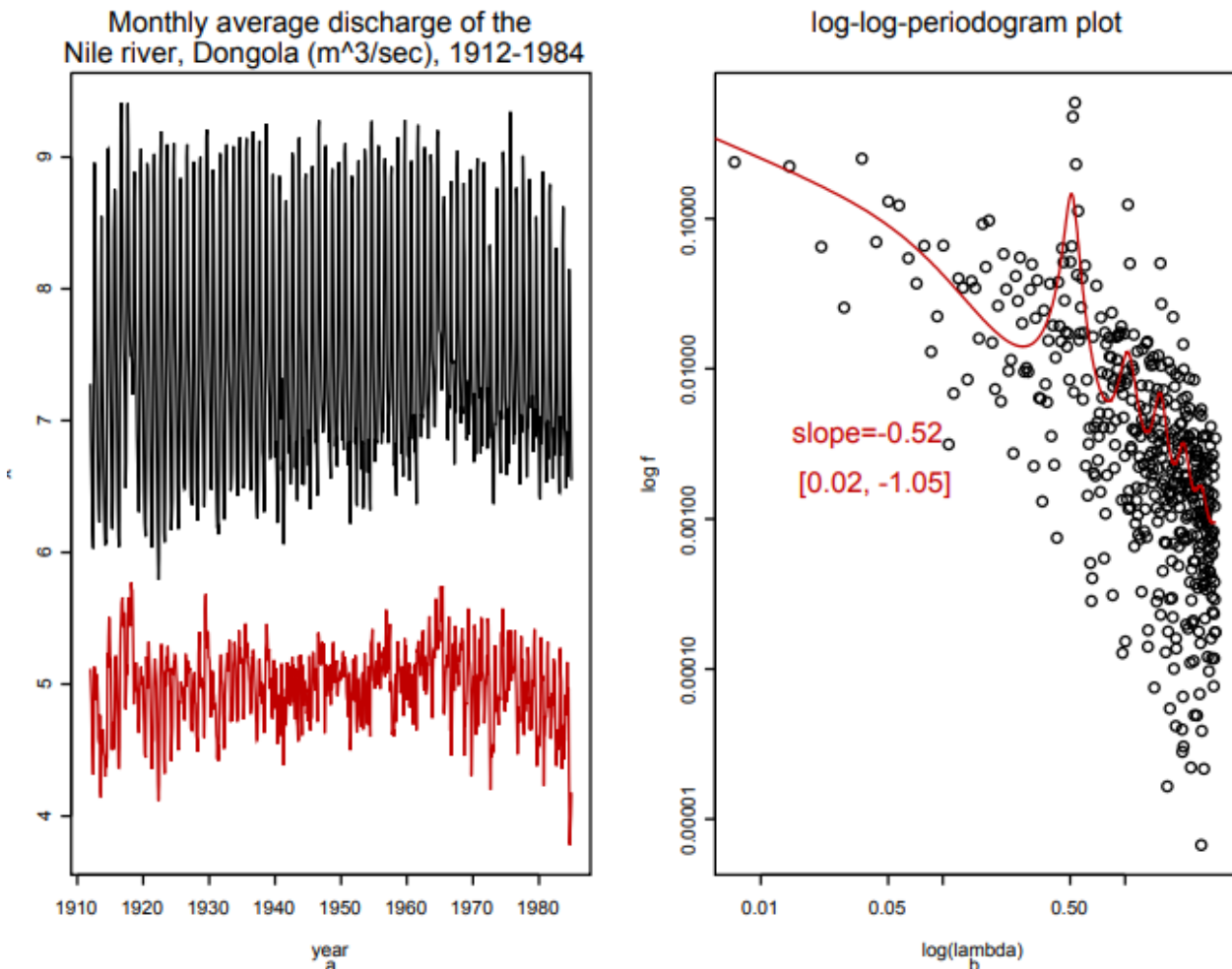


Рисунок 1.12 – Щомісячні вимірювання стоку Ніла

Можлива причина цього – зміна «сезонного ефекту» із плином часу, він може бути нелінійним, або стохастичним. Підсистема MLE в поєднанні з БІК фіксує рештки впливу сезонності досить добре. Ця модель припускає, що сезонність (що залишається після попереднього віднімання середніх значень) стохастична.

Розглянуті вище приклади можуть моделюватися за допомогою стаціонарних процесів, але часто стаціонарність не виглядає реалістичним

припущенням в таких випадках, або, як мінімум, не є гарантовано визначеною. Це ускладнює ідентифікацію стохастичної довгої пам'яті, оскільки типові риси довготривалих процесів можуть бути змішані з нестационарними компонентами. Однак, виявлення та оцінка можливих компонентів довготривалої пам'яті необхідна для правильного висновку про ці нестационарні компоненти.

Типовим прикладом можна вважати оцінку глобального потепління. Рисунок 1.13-а) зображає середньорічну температуру центральної Англії з 1659 до 2010 року, а головне питання чи доводять ці дослідження систематичне зростання місцевих температур. Найлегший спосіб відповісти на це питання побудувати пряму і визначити значення схилу β , а точніше чи додатне воно. Якщо спостереження задано $y_t = \beta_0 + \beta_1 t + e_t$, де e_t – стаціонарна, з довготривалою залежністю такою, що $p(k) \sim c|k|^{2d-1}, |k| \rightarrow \infty, d \in (0, \frac{1}{2})$, то дисперсія найменших квадратів статистичної оцінки β_1 збільшує в константний період часу n^{2d} у порівнянні з некорельованими або слабозалежними випадками.

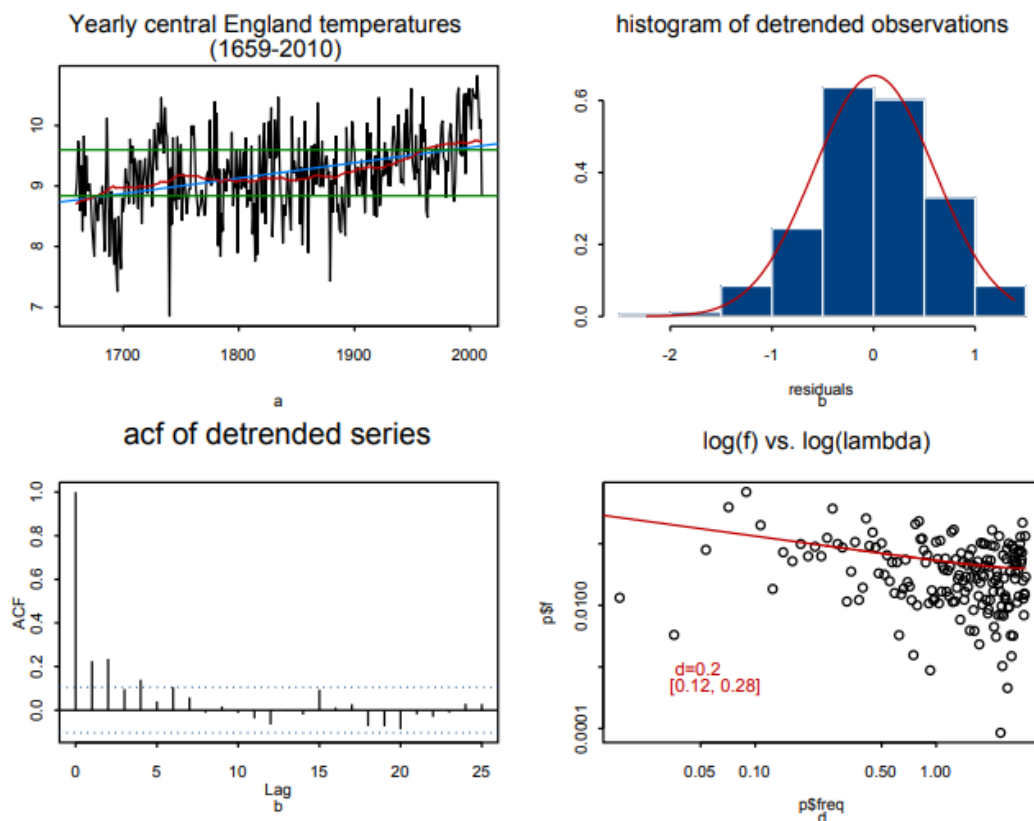


Рисунок 1.13 – Середньорічна температура центральної Англії

1.4 Адекватність моделі

Побудова коректної моделі часових рядів вимагає їх стаціонарності (або коінтеграції для багатовимірного випадку). У разі, якщо вихідні ряди не є стаціонарними, перед здійсненням оцінювання моделі необхідно застосувати процедуру приведення розглянутих рядів до стаціонарного виду. Для перевірки стаціонарності і визначення порядку інтеграції досліджуваних рядів буде застосовано кілька альтернативних тестів, що обумовлено важливістю гіпотези про стаціонарності рядів для подальшого моделювання.

Всі тести такого роду можна розділити на дві великі категорії в залежності від того, яка гіпотеза розглядається як нульовий. Розширений тест Діккея - Фуллера (ADF) і тест Філіпса - Перона (PP), що представляють собою одні з найбільш популярних і широко відомих інструментів аналізу поведінки часових рядів, перевіряють нульову гіпотезу про нестаціонарності процесу при альтернативній гіпотезі про те, що процес стаціонарний.

Тест Філіпса - Перона заснований на непараметричній корекції статистики Діккея - Фуллера, спрямованої на врахування можливої гетероскедастичності помилок, але це бачиться особливо актуальним саме в рамках аналізу фінансових ринків, так як фінансові часові ряди часто демонструють ефекти типу ARCH.

Результати цих двох тестів виявляють присутність одиничного кореня у всіх рядах валютних курсів без винятку, що говорить про їх нестаціонарності. Треба відзначити, що цей висновок узгоджується з переважною більшістю досліджень, присвячених фінансовим часовим рядам. Другим висновком, який може бути зроблений, виходячи з проведених розрахунків, є стаціонарність цих рядів в перших різницях або відсутність одиничних коренів в рядах доходностей.

Тестування статистичних гіпотез має на увазі, що нульова гіпотеза відкидається тільки при наявності на це дуже серйозних підстав.

У разі, якщо дослідник хоче використовувати стаціонарність як нульової гіпотези і має підстави для переваги саме цього припущення, зручність традиційних тестів, перерахованих вище, виявляється під питанням. Квятковський, Філіпс, Шмідт, Шин (KPSS) розробили альтернативний підхід для перевірки наявності одиничних коренів, що базується на стаціонарності ряду в якості нульової гіпотези.

Також існує ще один тест, який належить до цієї групи і орієнтований безпосередньо на тестування наявності довгої пам'яті в якості альтернативної гіпотези - тест Ло (1991). Теверовські (1999) зазначив, що ймовірність прийняття нульової гіпотези про відсутність довгострокової залежності завжди виявляється переоціненою цим тестом, тим не менш, Кирман і Тейссієр (2002) показали, що цей тест є більш потужним, ніж тест KPSS.

Комбінування результатів, отриманих на основі статистик Діккеля - Фуллера (і Філіпса - Перона) і KPSS (і Ло) призводить до чотирьох можливим наслідків:

- відкидання нульової гіпотези статистикою ADF і неможливість відкинути нульову гіпотезу статистикою KPSS розглядається як серйозне свідчення на користь стаціонарності ряду, тобто процесу $I(0)$;
- неможливість відкинути нульову гіпотезу тестом ADF і відкидання цієї гіпотези тестом KPSS служить важливим індикатором наявності одиничного кореня, тобто процесу $I(1)$;
- неможливість відкидання нульових гіпотез обома (ADF і KPSS) статистиками імовірно може свідчити про недостатній інформативності даних щодо довгострокових характеристик;
- одночасне відкидання нульових гіпотез статистиками ADF і KPSS вказує з високою часткою ймовірності на те, що даний процес не може бути описаний ні як процес $I(0)$, ні як $I(1)$.

Можна констатувати, що в нашому випадку ряди валютних курсів слідують, без сумніву, процесу $I(1)$, тоді як зробити остаточні висновки щодо рядів доходностей виявляється складніше. Тести ADF і PP однозначно відкидають гіпотезу про наявність поодиноких коренів у всіх рядах доходностей. Результати, отримані з використанням тесту KPSS, виявляються менш однозначними і чутливими до вибору модифікації цього тесту (з наявністю або відсутністю тренда). При цьому, як було вже сказано, тестові процедури надають перевагу нульовій гіпотезі, яка може бути відкинута тільки при серйозних свідоцтвах не на її користь. У зв'язку з цим представляється можливим, що для процесів $I(d)$ з невеликими величинами параметра d , потужність тесту KPSS виявляється недостатньою для поділу процесів $I(0)$ і $I(d)$. При цьому тест Ло чітко вказує на відкидання гіпотези про короткої пам'яті для доходностей трьох валютних курсів (французький франк, німецька марка і голландський гульден). Таким чином, можна говорити про високу ймовірність можливості опису трьох рядів доходностей процесом виду $I(d)$. В роботі, однак, буде зроблена спроба оцінки моделей з довгою пам'яттю і для інших валют, оскільки, в силу обмеженості сили наявних тестів, свідоцтва, що підтримують припущення про коректність опису інших рядів процесом $I(0)$, подаються не цілком достатніми.

Висновки до розділу та постановка задачі

Процеси з довгою пам'яттю вперше були досліджені інженером Херстом при вивченні приток Нілу і оптимальних розмірів резервуарів води. До початку 1950-х рр моделі, традиційно використовувалися в гідрології, припускали, що притоки і відтоки річок є незалежними в часі або, по крайній мере, слабо залежними. У цих умовах вивчення історії рівнів води не привносив ніякої корисної інформації для розуміння майбутнього поведінки річки. Іншими словами, ефект шоку, вироблений, наприклад, великим паводком, повинен був

швидко зникати, не впливаючи на досліджуваний ряд. В цьому випадку ми говоримо про нульовий пам'яті (повна залежність) або про короткої пам'яті (слабка залежність між спостереженнями і короткострокова залежність). Хьорст показав, що ці моделі в значній мірі недооцінюють складність гідрологічних коливань, а вивчення попередніх рівнів води надає вкрай цінну інформацію щодо майбутньої поведінки річки.

Необхідно відзначити, що поняття довгою пам'яті є проміжним між поняттями пам'яті короткою і пам'яті нескінченною. Це зауваження набуває особливої важливості, оскільки часто виявляється скрутним проведення жорсткого розмежування між цими типами поведінки тимчасових поруч, особливо в разі розгляду економічних даних. У випадку з короткою пам'яттю ефект шоку не впливає на поведінку аналізованого ряду в довгостроковому періоді. Навпаки, для ряду, що характеризується нескінченною пам'яттю, ефект шоку позначається на всіх без винятку майбутніх значеннях даного ряду. У проміжному випадку наявність в тимчасовому ряді ефекту довгострокової пам'яті призводить до вкрай тривалим, але не перманентним наслідків шоку. Ці наслідки не є постійними в тому сенсі, що рано чи пізно ряд повернеться до свого рівноважного рівня (такий рівень часто називають «нормальним» або «природним»), тоді як ряд з нескінченною пам'яттю ніколи не повертається до рівноважного рівня після події шоку.

Один із способів охарактеризувати короткотривалу і довготривалу залежність полягає у аналізі їх функцій автокореляцій. У процесах із короткою пам'яттю зв'язок між значеннями в різні періоди часу швидко зменшуються при зростанні різниці у часі вимірювання. Також автокореляція може зменшуватися до нуля, після досягання певного часового проміжку між вимірюваннями чи затухати експоненціально. Натомість, зв'язок між значеннями набагато сильніший, а затухання повільніше за експоненціальне.

Другий спосіб описується за допомогою дисперсії часткової суми послідовних значень: для процесів із короткою пам'яттю дисперсія випадкової

величини зростає пропорційно кількості виразів, а у процесів із довгою пам'яттю дисперсія збільшується швидше, та практично завжди являє собою експоненціальну функцію із експонентою більшою за одиницю. Спосіб визначення цієї поведінки використовує діапазон масштабування (rescaled range).

Описані вище методи математично пов'язані один з одним, але це не єдині способи визначення довготривалої залежності. У випадках, коли авоковаріації поцесу не існує, часто користуються допомогою с самоподібних процесів.

Параметр Херста (H) являє собою міру величини довготривалої залежності в часових рядах (хоча це має інше значення в контексті самоподібних процесів). Цей параметр сягає значень від 0 до 1, при значенні 0,5 довготривала залежність відсутня, але чим ближче значення H до 1, тим вища ступінь впливу двоготривалої залежності. H менше від 0,5 відповідає антиперсистенції, яка, на відміну від довготривалої залежності, вказує на сильну негативну кореляцію.

РОЗДІЛ 2. МЕТОДИКА МОДЕЛЮВАННЯ ПРОЦЕСІВ ІЗ ДОВГОЮ ПАМ'ЯТТЮ

2.1 Особливості структури загальних ймовірнісних моделей

Розглянемо деякі приклади процесів з довгою пам'яттю та моделі, що їх описують.

2.1.1 Гаусів процес

Гаусів процес має довгу пам'ять, якщо і тільки якщо його автоковаріаційна функція має вигляд $p(k) \sim c|k|^{2d-1}, |k| \rightarrow \infty, d \in (-\frac{1}{2}, \frac{1}{2})$ з показником Херста $2 - 2H, 1/2 < H < 1$.

Лінійний процес $x_t = \sum_{j \in \mathbb{Z}} a_j \varepsilon_{t-j}$, $\{\varepsilon_t\}$ - послідовність незалежних однаково розподілених випадкових величин. Якщо $\{\varepsilon_t\}$ - гаусів білий шум, тоді $\{x_t\}$ є гаусовим. Достатньою умовою довгої пам'яті є умова $a_j = l(j)j^{-d-1/2}, d \in (0, 1/2), l(j)$ - повільно мінлива функція на нескінченності. Тоді x_t володіє довгою пам'яттю з показником Херста $H = d + 1/2$.

2.1.2 Фрактальний гауссів шум

Цей процес визначається як $X_t = B_H(t) - B_H(t-1)$, $t \in \mathbb{R}$ і має автоковаріаційну функцію наступного вигляду:

$$\text{cov}(X_{s1} X_t) = \frac{1}{2} \left(\left| t-s \right|^{-1} \right)^{2H} + \left| t-s \right| + 1 \left| t-s \right|^{2H} - 2 \left| t-s \right|^{2H} \right) \sim H \left(H - \frac{1}{2} \right) \left| t-s \right|^{2(H-1)},$$

при $|t-s| \rightarrow \infty$.

Тоді процес x_t має параметр довгої пам'яті $d = H - \frac{1}{2}$.

2.1.3 Фрактальний броунівський рух

$$\text{cov}(B_H(s), B_H(t)) = \frac{1}{2} \left(|t|^{2H} + |s|^{2H} - 2|t-s|^{2H} \right).$$

2.1.4 Авторегресійна модель дробно-інтегрованого рухомого середнього

Математичною моделлю процесу з довгою пам'яттю є запропонована в $\varphi(L)(1-L)^d X_t = \theta(L)\eta_t$ лінійна авторегресійна дрібно-інтегрована модель рухомого середнього ARFIMA (p, d, q), є узагальненням лінійної авторегресійної інтегрованої моделі рухомого середнього ARIMA (p, q). Вважають, що процес x_t є авторегресійним порядку p проінтегрованим процесом $I(d)$ порядку d рухомого середнього порядку q , якщо має місце наступний його вигляд:

$$(1-L)^d y_t = \varepsilon_t, \quad t = 1, \dots, T,$$

$$\varphi(L)\varepsilon_t = \theta(L)\eta_t, \quad \phi(L)\varepsilon_t = \theta(L)\eta_t, \quad \eta_t \sim N(0, \sigma^2).$$

$$\varphi(L)(1-L)^d X_t = \theta(L)\eta_t,$$

де L – оператор зворотнього зсуву ($LX_t = X_{t-1}$), $\varphi(L) = 1 - \varphi_1 L - \dots - \varphi_p L^p$,

$\theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$. Оператор L називається лаговим оператором і процес ε_t є процесом з короткою пам'яттю, $I(0)$. Параметр d мається на увазі в цьому випадку дрібним числом з області $\left[-\frac{1}{2}, \frac{1}{2}\right]$. Оператор $(1-L)^d$ має наступне розкладання в ряд:

$$(1-B)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)} B^j,$$

де $\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt$ – гамма функція. Автокореляційна функція такого ряду

має вигляд

$$\rho_k = \frac{\Gamma(1-d)\Gamma(k+d)}{\Gamma(d)\Gamma(k+1-d)} \sim \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1}, \text{ при } k \rightarrow \infty.$$

Для $d \in (0, 1/2)$, де $c > 0$. В цьому випадку процес x_t є стаціонарним. При $d = 0$ процес має коротку пам'ять, по суті, це білий шум $I(d = 0)$. Випадок, у якому $d \in (0, 1/2)$ є достатньо рідкісним і описується як властивість антиперсистентності ряду. Для $d \in (0, 1/2)$ x_t є нестаціонарним процесом, що вимагає додаткових досліджень та обробки. При $d = 1$ процес x_t слідує процесам одиничного кореня. При $d \in (0, 3/2)$ процес є нестаціонарним і вимагається або його приведення до стаціонарного ряду або застосування спеціальних процедур оцінювання d , які є інваріантними до тренду.

2.2. Методи оцінки параметру довгої пам'яті

Серед різноманітних методів оцінки параметра d найбільше поширення отримали два підходи: методи періодограми і локальної оцінки Уїттла. Періодограмою часового ряду $\{x_t\}, t = 1, \dots, T$ називається

$$I_i = I(\lambda_i) = \frac{1}{2\pi T} \left| \sum_{t=1}^T x_t e^{it\lambda_i} \right|^2,$$

де $\lambda_i = 2\pi i / T$, $1 \leq i < m$, де m – число ортант періодограми, яке включене в періодограму таким чином, що воно зростає повільніше, ніж розмірність виборки $m / T \rightarrow 0$. Оцінка Портера–Худака базується на оцінці метода найменших квадратів рівняння лінійної логарифмічної регресії наступного вигляду:

$$\log(I(\lambda_i)) \rightarrow c + d_{LP} R_i + \varepsilon_i,$$

$$R_i = \log(4 \sin^2(\lambda_i / 2)).$$

Для оцінки $R(d)$ справедливо:

$$\sqrt{m} (\hat{d}_{LP} - d) \sim N\left(0, \frac{\pi^2}{24}\right).$$

Стандартна похибка оцінки складає $\pi / \sqrt{24m}$. Оптимальний вибір значення m складає:

$$m_{LP} = \left(\frac{27}{512\pi^2} \right)^{1/5} \frac{1}{\tau^{2/5}} T^{4/5}.$$

Метод локальної оцінки Уїтгла базується на мінімізації логарифмічної функції максимальної правдоподібності у спектральній області наступного вигляду:

$$R(d) = \log(G(d)) - \frac{2d}{m} \sum_{j=1}^m \log(\lambda_j),$$

$$G(d) = \frac{1}{m} \sum_{j=1}^m \lambda_j^{2d} I_y(\lambda_j).$$

Кюнч запропонував локальну оцінку Уїтгла, у який підсумування проводиться не по всій області частот. Замість цього вважається, що $m \rightarrow \infty$ трохи повільніше, ніж розміреність виборки $T \rightarrow \infty$ таким чином, що

$$\frac{1}{m} + \frac{m}{T} \rightarrow 0, \text{ при } T \rightarrow \infty.$$

Тоді локальна оцінка Уїтгла має наступний вигляд

$$\hat{d}_{LW} = \arg \min_{d \in [d_1, d_2]} R(d).$$

Робінсон довів, що при визначених умов справедливо

$$\mathfrak{R}_d = \sqrt{m} (\hat{d}_{LW} - d) \rightarrow N\left(0, \frac{1}{4}\right),$$

$$\text{для } \left[-\frac{1}{2}, \frac{1}{2} \right].$$

Даний результат може бути розширений до $d < 0,75$, більш того, локальна оцінка Уїттла залишається конзистентною і при $d < 1$. Оцінка є доволі розповсюдженою на практиці, оскільки вона більш ефективна, ніж метод оцінювання GPH. Практичне застосування розробленого методу впирається в оптимальний вибір значення ширини спектральної області підсумування, тобто величини m , вибір якої визначається відношенням варіації оцінки $\hat{m}^{(k+1)}$ та її зміщенням.

Зупинимось на цій проблемі більш ретельно. Як відомо, вибір занадто великої або надто малої області несе по собі надто велику зміщеність і, відповідно, дуже велику варіацію оцінки Уїттла. Так, наприклад, для знаходження балансу у відношенні даних показників оптимальне значення m пропонується обирати в наступному вигляді

$$m = \left(\frac{3}{4\pi} \right)^{4/5} \left(\frac{1}{\tau + d/12} \right)^{2/5} T^{4/5}.$$

Не дивлячись на те, що, як правило, кількість гармонік періодограми обирається фіксована, в роботі пропонується ітераційний алгоритм знаходження оптимального значення m , що базується на методі мінімізації середньоквадратичної помилки оцінки:

$$\hat{d}^{(k)} = \arg \min R(\hat{m}^{(k)}, d),$$

$$\hat{m}^{(k+1)} = \left(\frac{3}{4\pi} \right)^{4/5} \left| \theta + \frac{\hat{d}^{(k)}}{12} \right|^{-2/5} T^{4/5},$$

Для $\begin{bmatrix} -\frac{1}{2}, \frac{1}{2} \end{bmatrix}$ і початкового значення $\hat{m}^{(0)} = T^{\frac{4}{5}}$. У випадку моделі ARFIMA

параметр θ може бути наближено обраним як $\theta = \frac{f_x''(0)}{2f_x'(0)}$, де $f_x(\lambda)$ спектральна щільність ARMA процесу. Також в даній роботі показано, що ефективна оцінка для цього параметра може бути на оцінці регресії методом найменших квадратів:

$$I(\lambda_j) = \sum_{k=0}^2 Z_{jk}(\hat{d}^{(0)}) \tilde{\varphi}_k + \tilde{\varepsilon}_k, \quad Z_{jk}(d) = \frac{|1 - \exp(i\lambda_j)|^{-2d} \lambda_j^k}{k!},$$

$$I(\lambda_j) = |1 - \exp(i\lambda_j)|^{-2\hat{d}^{(0)}} \left(\tilde{\varphi}_0 + \tilde{\varphi}_1 \lambda_j + \tilde{\varphi}_2 \frac{\lambda_j^2}{2} \right) + \tilde{\varepsilon}_j, \quad j = 1, \dots, \hat{m}^{(0)}.$$

Оцінками $f'(0)$ и $f_x''(0) \in \tilde{\varphi}_0$ и $\tilde{\varphi}_2$ відповідно, отже величину θ в можна замінити на $\hat{\theta} = \tilde{\varphi}_2 / 2\tilde{\varphi}_0$.

Суттєвою проблемою при оцінці параметра d є визначення типу стаціонарності досліджуваних даних. Як правило, для таких цілей застосовується широкий клас відомих критеріїв на стаціонарність і одиничні корені. Однак, їх потужності не завжди достатньо для того, щоб з високою надійністю виявляти наявність дрібно-інтегрованих процесів і, відповідно, довготривалої пам'яті. Стосовно $I(d)$ процесів проблема полягає у виборі $d \geq 1/2$ чи $d < 1/2$, так як $I(d)$ є нестаціонарними при $d \geq 1/2$.

У роботі Henry M.(2001) «Robust automatic bandwidth for long memory. Journal of Time Series Analysis, 22: 293—316.» запропонований новий метод оцінки d , так званий метод точної локальної оцінки Уїттла (ELW), який дозволяє отримати оцінку d в незалежності від області стаціонарності процесу. ELW-оцінка є конзистентною та має нормальний розподіл $N(0; 1/4)$ для всіх значень

d , якщо процедура оцінювання поєднує інтервал значень d не менш ніж $9/2$ і відомо математичне сподівання.

Однак, як правило, математичне сподівання економічних часових рядів невідомо і являє собою поліноміальні тренди, яке досить складно виділити. В результаті досліджень доведено: по-перше, якщо невідоме математичне сподівання замінити на середнє вибірки, ELW-оцінка конзистентна при $d > 0$ і асимптотично нормальна для $d \in (0,2)$, але при цьому необхідно накласти більш жорсткі обмеження на кількість використаних компонент періодограми, що використовуються при оцінці.

В цій роботі запропонований удосконалений двокроковий метод локальної точної оцінки Уїтла, що комбінує у собі обидва ці підходи. Цільова функція модифікована таким чином, щоб вона поєднувала у собі як медіанну оцінку вибірки, так і оцінку початкового значення, залежну від d .

Модифікований таким чином ELW-метод, має граничне нормальний розподіл $N(0, 1/4)$, а одержувані з його допомогою оцінки є конзистентними для $d > -1/2$.

2.3 Методика моделювання процесів із довгою пам'яттю

Процесу з довгою пам'яттю можна дати і цілком чітке визначення, формалізує описані вище його особливості. Стаціонарний процес X_t є процесом з довгою пам'яттю, якщо існує дійсне число a , $0 < a < 1$, і константа c , $c > 0$, для яких виконується умова:

$$\lim_{k \rightarrow \infty} \frac{\rho_k}{c^* k^{-a}} = 1 ,$$

де ρ – функція автокореляції, та k – номер лага.

Тобто автокорреляції процесу з довгою пам'яттю задовольняють наступній асимптотичному співвідношенню: $\rho_k \sim c^* k^{-\alpha}$ при $k \rightarrow \infty$. Таким чином, спадання автокореляцій відбувається вкрай повільно, по гіперболі. Це гіперболічне спадання автокореляцій протиставляється їх експоненціального зменшенням, характерному для процесів з короткою пам'яттю (ARMA): $\rho_k \leq c^* a^k$ де c – додатна константа і $0 < a < 1$.

Важливо відзначити, що визначення довгої пам'яті є асимптотичним. Іншими словами, розглядаючи автокорреляції, можна констатувати, що кожна з автокореляцій окремо може бути слабкою, тоді як їх сума повинна бути високою. Це робить ще більш складним виявлення довгою пам'яті, оскільки виявлення окремих лагов, для яких автокорреляція є високою, не виявляється достатнім. В даному випадку важливим виявляється темп прагнення автокореляцій до нуля в міру збільшення лага. Чим вище ступінь персистентності процесу, тим повільнішою буде конвергенція до нуля і тим більшою буде сума кореляцій. Навпаки, для процесу з короткою пам'яттю сума автокореляцій буде менш значною, оскільки ці процеси зазвичай характеризуються однією або двома значущими автокореляціями при малих лагах.

2.3.1 Процес ARIFMA

Будь-які прикладні дослідження вимагають більш суворого опису всього поведінки ряду, яке може бути отримано на основі використання процесів ARFIMA.

У роботах, присвячених аналізу часових рядів, першим кроком є, як правило, визначення порядку інтегрованості ряду (мова йде про значення параметра d процесу ARIMA (p, d, q)). Як правило, автори обмежуються вибором

між d рівним 0 і 1. Випадок $d = 0$ відповідає короткій пам'яті ряду, тоді як при $d = 1$ можна зробити висновок про нескінченну пам'яті. Говорячи про нескінченну пам'яті, мають на увазі, що кожен шок впливає на поведінку ряду нескінченно довго.

З іншого боку, існування короткої пам'яті означає, що наслідки шоку зникнуть досить швидко, при цьому з поля зору дослідників вислизає проміжна ситуація, коли наслідки шоку виявляються тимчасовими, але тривалими (тобто випадок довгої пам'яті, або персистентності). Для вирішення цієї проблеми Гренджер і Джой (1980) запропонували новий клас моделей ARFIMA (p, d, q), що опускає можливість нецілого параметра d .

Кажуть, що ряд x_t слідує процесу ARIFMA (p, d, q), якщо:

$$\Psi(L)(1-L)^d X_t = \Theta(L)\varepsilon_t$$

$$(1-L)^d = 1 - dL - \frac{d(1-d)}{2!}L^2 - \frac{d(1-d)(2-d)}{3!}L^3 - \dots =$$

$$= \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)} L^k = 1 - \sum_{k=1}^{\infty} c_k(d) L^k$$

Процеси ARMA (p, q) і ARIMA (p, d, q) є випадками цього процесу при $d = 0$ і $d = 1$ відповідно.

Процеси ARFIMA є вкрай зручний інструмент аналізу часових рядів, оскільки дають можливість одночасного моделювання ефектів довгої і короткої пам'яті. Моделювання довгою пам'яті на основі використання традиційних процесів ARMA з включенням вкрай великої кількості лагів також можливо (автокорреляція буде являти собою в цьому випадку суму експонент), проте таке моделювання не дозволяє оцінити ступінь персистентності ряду і, крім того, в цьому випадку для персистентного ряду ми отримуємо коріння поліномів

авторегресії змінного середнього близькими до одиниці, внаслідок чого асимптотические розподілу виявляються поганими апроксимаціями розподілів на кінцеві вибірки.

У своєму дослідженні, присвяченому порівнянню моделювання стаціонарних часових рядів за допомогою процесів ARMA і ARFIMA, Андерсон (1998), застосовуючи симуляції Монте-Карло і порівнюючи помилки прогнозів, показує, що ігнорування довгою пам'яті, коли вона в дійсності має місце, призводить до більш серйозного погіршення результатів, ніж її накладення при відсутності такої.

Це спостереження є вкрай важливим, оскільки на практиці досліднику ніколи не відомо, який процес в дійсності лежить в основі динаміки цін фінансових активів. На основі наведених вище міркувань можна розглядати використання процесів ARFIMA як один з найбільш сучасних і актуальних підходів для вивчення фінансових часових рядів.

Наявність довгої пам'яті означає, що інформація, що міститься в минулому поведінці ряду, виявляється корисною для передбачення його майбутніх значень і дає можливість зробити спробу перевершити прогноз на основі процесу "випадкового блукання".

2.3.2 Процеси із довгою пам'яттю в дисперсіях

Описуючи моделі ARFIMA, ми припускали, що ε_t є "білим шумом" і має дисперсію, постійну в часі. Однак емпіричні фінансові дані показують, що волатильність фінансових рядів помітно змінюється в часі. Для моделювання цього властивості фінансових ринків дослідники, як правило, вдаються до використання моделей сімейства GARCH, що дозволяють оцінити можливу

залежність поточної волатильності від її попередніх значень і квадратів величин попередніх помилок.

Процеси FIGARCH. Почнемо з розгляду найбільш традиційної моделі GARCH, запропонованої Боллерслевом. Формально модель визначається як

$$\sigma_t^2 = \omega + \alpha(L)\varepsilon_t^2 + \beta(L)\sigma_t^2,$$

де L – лаговий оператор $\alpha(L) \equiv \alpha_1 L + \alpha_2 L^2 + \dots + \alpha_r L^r$ і $\beta(L) \equiv \beta_1 L + \beta_2 L^2 + \dots + \beta_s L^s$.
Процес GARCH (r,s) може бути також представлений у вигляді процесу ARCH нескінченного порядку:

$$\sigma_t^2 = \omega[1 - \beta(L)]^{-1} + \alpha(L)[1 - \beta(L)]^{-1} \varepsilon_t^2 \equiv \omega[1 - \beta(L)]^{-1} + \lambda(L)\varepsilon_t^2.$$

Цей процес може бути представлений і як процес ARMA (m, r) щодо:

$$[1 - \alpha(L) - \beta(L)]\varepsilon_t^2 = \omega + [1 - \beta(L)]v_t,$$

$$\text{де } m = \max\{r, s\}, v_t \equiv \varepsilon_t^2 - \sigma_t^2.$$

Можна інтерпретувати некоррелірований процес з ненульовим середнім як "інновації" в умовній дисперсії. У разі, коли Лагові поліном авторегресії містить одиничний корінь, процес називають інтегрованим в дисперсії. Відповідний клас моделей називають інтегрованими GARCH (p, q) моделями (IGARCH) і описують рівнянням:

$$\begin{aligned} \phi(L)(1-L)\varepsilon_t^2 &= \omega + [1 - \beta(L)]v_t, \\ \phi(L) &\equiv [1 - \alpha(L) - \beta(L)](1-L)^{-1} \end{aligned}$$

Боллерслев і Енгл (1993) визначають процес як інтегрований в дисперсіях, якщо $\lim_{j \rightarrow \infty} E_t [\text{var}_{t+j-1}(y_{t+j})]$ не сходиться до деякої константи з одиничною ймовірністю для кожного періоду t , тобто довгострокові прогнози умовної дисперсії залишаються чутливими до початкових умов на всіх горизонтах прогнозу. Хоча емпіричні оцінки часто не дозволяють відкинути гіпотезу про наявність одиничного кореня в дисперсії, інтуїція підказує, що в більшості випадків дисперсія повинна повертатися до середнього значення. Узгодження цих двох точок зору можливо при використанні, як і в моделі ARFIMA, часткового порядку інтегрування.

Частково інтегрований процес GARCH (Або FIGARCH (r, x, s)), може бути отриманий шляхом заміни в моделі IGARCH оператора перших різниць на оператор часткового диференціювання:

$$\phi(L)(1-L)^s \varepsilon_t^2 = \omega + [1 - \beta(L)]v_t.$$

Очевидно, що модель FIGARCH (r, x, s) зводиться до моделі GARCH (r, s) при $x = 0$ і до моделі IGARCH (r, s) при $x = 1$. Можливість використання значень x в інтервалі від нуля до одиниці надає моделі додаткову гнучкість, яка може бути необхідною при моделюванні довгострокових залежностей, властивих умовної дисперсії фінансових часових рядів. Можна провести аналогію між цим класом моделей і моделями ARFIMA для умовного середнього. Як і в випадку процесів ARFIMA, короткострокова динаміка умовної дисперсії описується традиційними параметрами GARCH, тоді як довгострокові залежності описуються параметром часткового диференціювання.

Процеси FIAPARCH. У попередньому розділі було згадано, що "інновації" можуть надавати асиметричне вплив на волатильність. Для обліку цього емпіричного спостереження Дінг, Гренджер і Енгл (1993) запропонували модель APARCH (r, s) (Asymetric Power ARCH). Ця модель може бути представлена як:

$$\sigma_t^\delta = \omega + \sum_{i=1}^s (\alpha_i |\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta + \sum_{j=1}^r \beta_j \sigma_{t-j}^\delta,$$

де $\delta > 0$ і $-1 < \gamma_i < 1$ ($i = 1, \dots, s$). Автори також звертають увагу на важливість статечного параметра, з використанням якого відбувається перетворення ряду.

Не існує будь-яких підстав стверджувати, що феномен кластеризації волатильності є властивістю виключно квадратів прибутковості. Взагалі абсолютна зміна цін фінансових активів також демонструватиме аналогічну кластеризацію волатильності, і використання статечного параметра проводиться для того, щоб підкреслити відносно спокійні і бурхливі періоди.

Дійсно, може бути застосований будь-який позитивний статечної параметр, і стандартне використання в цій ролі квадратичних величин представляється більшою мірою відображенням гіпотези нормальності розподілу даних, до якої традиційно прийнято звертатися.

У разі, якщо ряд нормально розподілений, то він може бути в повній мірі охарактеризований першими двома моментами і, отже, виправданим є включення в модель квадратичної величини, що представляє собою міру дисперсії.

Однак, в разі прийняття допущення про те, що розподіл даних може бути ненормальним, для адекватного опису ряду виникає необхідність піднятися в область моментів більш високого порядку (асиметрія, ексцес і т.д.). У цьому випадку перевага квадратичного параметра втрачається і більш придатними можуть виявитися вже й інші поважні перетворення.

Можна припустити, що для негауссовских даних, шляхом зведення доходностей в квадрат, екзогенно накладається структура, яка потенційно може вести до субоптимальних результатів для моделювання і зниження ефективності

прогнозування в порівнянні з іншими статечними параметрами. Замість накладення обмежень на параметр ступеня, дана модель пропонує зробити його ендогенним і оцінити його оптимальну величину. У той же час ця модель дозволяє врахувати можливу асиметрію в реакції волатильності на шоки різних знаків. Таким чином гнучкість змінюється статечного параметра об'єднується з наявністю коефіцієнта асиметрії, і крім того, модель APARCH включає цілий ряд інших варіантів моделі ARCH як окремих випадків.

Згідно Дінгу, Гренджеру і Енглу (1993) при $w > 0$ і $\sum_{i=1}^s \alpha_i E(|z| - \gamma_i z)^\delta + \sum_{j=1}^r \beta_j < 1$ стаціонарне рішення рівняння описаного вище існує і приймає наступний вигляд:

$$E(\sigma_t^\delta) = \frac{\alpha_0}{1 - \sum_{i=1}^s \alpha_i (|z| - \gamma_i z)^\delta - \sum_{j=1}^r \beta_j}.$$

Потрібно відзначити, що в разі, якщо $\gamma = 0$, $\delta = 2$, $p = 1$, $q = 1$ і z має нульове середнє і одиничну дисперсію, ми отримуємо стандартні умови стаціонарності моделі GARCH (1,1) – $\alpha_1 + \beta_1 < 1$. При цьому, якщо $\gamma \neq 0$ і / або $\delta \neq 2$, дана умова залежить від гіпотез, зроблених щодо процесу, якого дотримуються "інновації".

Екстраполяція моделі на випадок фракційної інтеграції була здійснена Тсе (1998) і може бути представлена як:

$$\sigma_t^\delta = \omega + \left\{ 1 - [1 - \beta(L)]^{-1} \phi(L)(1-L)^\xi \right\} (|\varepsilon_t| - \gamma \varepsilon_t)^\delta.$$

2.3.3 Процеси із подвійною довгою пам'яттю

У даній роботі будуть об'єднані два описаних вище підходу і побудована модель, яка припускає присутність довгою пам'яті в значеннях ряду і його волатильності одночасно. Формально ця модель може бути визначена наступним чином:

$$\Phi(L)(1-L)^d(y_t - \mu) = \Theta(L)\varepsilon_t$$

$$\varepsilon_t = \sigma_t z_t$$

$$z_t \sim \text{i.i.d. } D(0, 1)$$

$$\sigma_t^2(\omega, \varphi, \beta, \xi) = \omega + \left[1 - (1 - \beta(L))^{-1} \varphi(L)(1-L)^\xi \right] \varepsilon_t^2,$$

де $D(\cdot)$ – функція щільності розподілення.

Важливо відзначити, що модель ARFIMA - FIGARCH є більше ніж простою комбінацією двох процесів з довгою пам'яттю. Їх спільне оцінювання є важливим для висновків щодо двох фракційних параметрів в значеннях ряду і волатильності. Інтуїція підказує наявність можливих паралелей з роботами, присвяченими тестування одиничних коренів в присутності структури типу ARCH в помилках, що дозволяє зробити висновок про те, що більш правильний опис динаміки волатильності може відігравати важливу роль в оцінці параметра часткової диференціації в рівнянні умовного середнього. Гренджер і Терасвірта (1993) показали, що помилки в специфікації рівняння для самих значень ряду також можуть призводити до зсувів в оцінках параметрів в рівнянні умовної дисперсії.

Аналогічним чином може бути побудована модель ARFIMA-FIAPARCH, шляхом заміни рівняння умовної волатильності в наведеній вище системі рівнянь. Такий вид моделювання пропонується в даній роботі вперше і не зустрічався раніше в літературі. В емпіричній частині роботи буде показана можливість поліпшення результатів оцінок з використанням моделей з подвійною довгою пам'яттю в порівнянні з оцінками, отриманими за допомогою

широкого спектру альтернативних моделей аналогічного класу, тобто сімейства ARMA, ARCH і їх різноманітних комбінацій і модифікацій.

2.3.4 Функції розподілу

У попередньому розділі було показано, що розподіл фінансових рядів часто має властивості, що дозволяють прийти до висновку про те, що воно не є нормальним. Зокрема, з огляду на згадувану вже актуальність теми "товстих хвостів" розподілів в емпіричних фінансах, можна очікувати, що використання більш відповідного цьому стилізованому фактом розподілу дозволить зменшити надмірну ексцес в залишках моделей з умовної гетероскедастичності. У зв'язку з цим багато авторів вдаються до використання розподілів з "товстими хвостами" в моделюванні фінансових часових рядів. Одним з найбільш часто зустрічаються розподілів такого виду є розподіл Стюдента.

Головним недоліком цього розподілу є те, що воно залишається симетричним, хоча і може враховувати властивість лептокуртічності ("товстих хвостів"). Для вирішення цієї проблеми в своїх недавніх роботах Ламбер і Лоран (2001) і Петерс (2001) успішно застосували асиметричну версію розподілу Стюдента в рамках моделі GARCH.

У зв'язку з тим, що дійсний розподіл тимчасового ряду ніколи невідомий, в наступному розділі, присвяченому емпіричному моделюванню, будуть використані три різних розподіли для порівняння отриманих за їх допомогою результатів і вибору найбільш відповідного варіанту.

Висновки до розділу

В даному розділі були розглянуті деякі приклади процесів з довгою пам'яттю та моделі, що їх описують. За допомогою цих моделей у наступному розділі будуть використані три різних розподіли, а за результатами експериментів прогнозування ми зможемо обрати найбільш відповідний варіант.

РОЗДІЛ 3. ВИКОНАННЯ ОБЧИСЛЮВАЛЬНИХ ЕКСПЕРИМЕНТІВ

3.1 Короткий опис системи R

R — мова програмування і програмне середовище для статистичних обчислень, аналізу та зображення даних в графічному вигляді. Розробка R відбувалась під істотним впливом двох наявних мов програмування: мови програмування S з семантикою успадкованою від Scheme. R названа за першою літерою імен її засновників Роса Іхаки (Ross Ihaka) та Роберта Джентлмена (Robert Gentleman) працівників Оклендського Університету в Новій Зеландії. Незважаючи на деякі принципові відмінності, більшість програм, написаних мовою програмування S запускаються в середовищі R.

R поширюється безкоштовно за ліцензією GNU General Public License у вигляді вільнодоступного вихідного коду або відкомпільованих бінарних версій більшості операційних систем: Linux, FreeBSD, Microsoft Windows, Mac OS X, Solaris. R використовує текстовий інтерфейс, однак існують різні графічні інтерфейси користувача (див. Графічні Редактори Скриптів та IDE).

R має значні можливості для здійснення статистичних аналізів, включаючи лінійну і нелінійну регресію, класичні статистичні тести, аналіз часових рядів (серій), кластерний аналіз і багато іншого. R легко розбудовується завдяки використанню додаткових функцій і пакетів доступних на сайті Comprehensive R Archive Network (CRAN). Більша частина стандартних функцій R, написана мовою R, однак існує можливість підключати код написаний C, C++, або Фортраном. Також за допомогою програмного коду на C або Java можна безпосередньо маніпулювати R об'єктами.

Хоча R орієнтована на розв'язок і аналіз статистичних задач, вона може використовуватися для матричних обрахунків з порівняльною швидкістю до математичних пакетів GNU Octave або MATLAB.

Створено багато пакетів для статистичних обчислень, біоінформатики, оптимізації тощо (див. "Пакети/Бібліотеки").

Середовище R містить засоби для візуалізації результатів обчислень (двовимірні, тривимірні графіки, діаграми, гістограми, діаграми (схеми) Ганта тощо). Графічні можливості R дозволяють створювати високоякісні графіки з різними атрибутами, зокрема математичні формули і символи.

Іншою особливістю є функція Sweave яка дозволяє інтеграцію і виконання коду R в документах написаних за допомогою LaTeX з метою створення динамічних звітів.

R de-facto став стандартом у міжнародній спільноті спеціалістів в галузі статистики, і широко використовується в розробках статистичних програм та аналізі даних. Згідно щорічному опитуванню Rexter's Annual Data Miner Survey в 2010 році, більшість (43%) серед опитаних спеціалістів з аналізу даних використовують у своїй роботі середовище R.

R належить до інтерпретованих мов програмування і для роботи використовується командний інтерпретатор. R підтримує концепцію Об'єктно-орієнтованого програмування (ООП) включаючи generic функції, результат виконання якої залежить від аргументів (типу об'єктів), що передаються generic функції. В мові програмування R всі змінні є об'єктами, кожен об'єкт належить до певного класу.

3.2 Приклади побудови моделі

В даному розділі буде розглянуто поведінку доходностей курсів валют ряду країн щодо долара США. В ході дослідження буде піддана перевірці гіпотеза про можливу наявність у даних рядах довгою пам'яті, будуть застосовані на практиці моделі і методи, описані в попередній частині роботи, з метою

вибору моделей, найбільш повно і адекватно описують реалії сучасних валютних ринків різних країн.

В роботі використані денні дані для обмінних курсів ряду валют (французький франк (FRF), англійський фунт стерлінгів (GBP), німецька марка (DEM), голландський гульден (NLG), японська ієна (YEN), канадський долар (CAD), швейцарський франк (CHF)) до долара США за період с 2 січня 1980 по 31 грудня 1998 року, тобто до введення євро (4774 спостереження) для FRF, DEM, NLG і з 2 січня 1980 по 30 липня 2002 року для інших валют. Обмінні курси виражені в одиницях даної валюти за долар США.

Для отримання рядів доходностей беруться логарифми вихідних рядів обмінних курсів, до них застосовується оператор перших різниць і отримані величини множаться на 100 для переходу до процентних змін і зміни масштабу.

$$r_t = 100 \cdot \ln \frac{P_t}{P_{t-1}}$$
 де r - прибутковість фінансового активу (в нашому випадку - валюти) за період від моменту $(t - 1)$ до t , P - обмінний курс в момент t . Основні описові статистики розглянутих рядів представлені в табл. 3.1.

Таблиця 3.1 – Описові статистики

Ряд	Середнє	Сер. відх.	Асиметрія	Ексцес	Харке-Бера
FRF	0,0069	0,6765	0,0540	6,6394	2636,5
DEM	-0,0007	0,6893	-0,1355	4,9224	749,6
NLG	-0,0003	0,6815	-0,1429	4,9719	789,6
GBP	0,0060	0,6349	0,0413	6,1078	2292,2
JPY	-0,0127	0,6929	-0,4981	6,6478	3391,3
CHF	-0,0016	0,7684	0,1125	14,3363	30490,8
CAD	0,0053	0,2868	0,0343	6,0960	2274,6

Тест Харке - Бера полягає в тестуванні нульової гіпотези про нормальність розподілу ряду. У разі, якщо прибутковості підкоряються нормальному закону, то статистика Харке - Бера слідує розподілу χ^2 - квадрат з двома ступенями свободи. Таким чином, якщо отримана статистика $JB \geq \chi^2_2$, нульова гіпотеза про нормальність відхиляється. Очевидно, що ця гіпотеза виявляється відкинutoю для всіх розглянутих рядів, оскільки отримані статистики виявляються значно вище критичних значень при будь-якому розумному рівні значущості. Всі ряди доходностей демонструють властивість лептокуртічності, при цьому параметри ексцесу виявляються близькими для різних рядів, але цікаво відзначити, що величина значень цієї характеристики розподілу значно нижче, ніж для біржових індексів (див. Мінйон (1998)). Таким чином, можна констатувати, що для валютного ринку менш характерні різкі скачки цін активів (в порівнянні з фондовим ринком), що може відображати як внутрішні особливості самого валютного ринку, так і роль центральних банків, інтервенції яких обмежують ступінь коливань валютних курсів. Коефіцієнти асиметрії мають різні знаки, проте помітно, що абсолютні величини цього параметра помітно вище для негативної асиметрії, хоча і в цьому випадку тільки коефіцієнт, отриманий для випадку японської ієни, можна порівняти з асиметрією біржових індексів (див. Мінйон (1998)). Негауссовських щільність розподілу розглянутих рядів показана на рис. 3.1, де безперервна лінія позначає оцінку щільності розподілу ряду, а пунктирна лінія - щільність нормального розподілу.

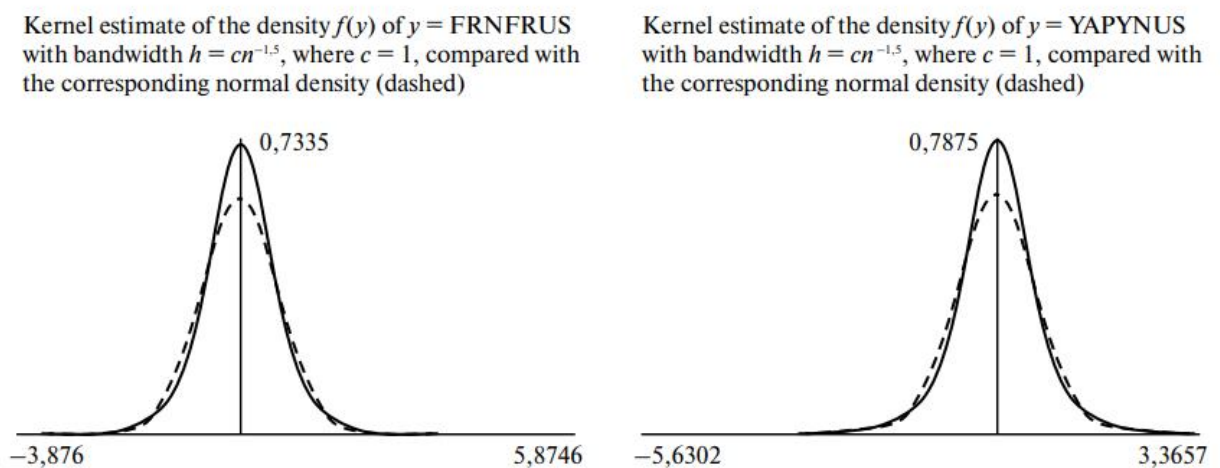


Рисунок 3.1 – Щільність розподілу FRF та YEN

Дані табл. 3.1 також вказують на стабільний характер величини стандартних відхилень для різних валютних курсів за винятком канадського долара. Для останнього валютного курсу був зафіксований значно нижчий рівень волатильності, що може бути пояснено наявністю тісних зв'язків між американською і канадською економіками і, як наслідок, тим, що курси валют цих країн щодо інших валют часто рухаються синхронно, тоді як коливання курсу американського долара до канадському виявляються досить обмеженими. Денна дисперсія для всіх інших валютних курсів знаходиться в межах від 0,665 до 0,763, що має на увазі середньорічну волатильність від 10,6 до 12,1%. Середні значення денних доходностей для різних валютних курсів мають різні знаки. Необхідно пояснити, що використані курси мають в якості базової валюти американський долар і, отже, позитивна прибутковість позначає здешевлення даної валюти щодо долара США і навпаки. Дані табл. 1 вказують на помітне подорожчання японської ієни (в середньому на 3,98% в рік) і швейцарського франка (в середньому на 0,78% в рік) і здешевлення французького франка (в середньому на 1,74% в рік), англійського фунта стерлінгів (1,51%) і канадського долара (1,46%). Німецька марка і голландський гульден демонструють тенденцію до деякого зміцнення, однак вона виявляється вкрай незначною.

Побудова коректної моделі часових рядів вимагає їх стаціонарності (або коінтеграції для багатовимірного випадку). У разі, якщо вихідні ряди не є стаціонарними, перед здійсненням оцінювання моделі необхідно застосувати процедуру приведення розглянутих рядів до стаціонарного виду. Для перевірки стаціонарності і визначення порядку інтеграції досліджуваних рядів буде застосовано кілька альтернативних тестів, що обумовлено важливістю гіпотези про стаціонарності рядів для подальшого моделювання.

Всі тести такого роду можна розділити на дві великі категорії в залежності від того, яка гіпотеза розглядається як нульовий. Розширений тест Діккеля - Фуллера (ADF) і тест Філіпса - Перона (PP), що представляють собою одні з найбільш популярних і широко відомих інструментів аналізу поведінки часових

рядів, перевіряють нульову гіпотезу про нестационарності процесу при альтернативній гіпотезі про те, що процес стаціонарний.

Тест Філіпса - Перона заснований на непараметричній корекції статистики Діккеля - Фуллера, спрямованої на врахування можливої гетероскедастичності помилок, але це бачиться особливо актуальним саме в рамках аналізу фінансових ринків, так як фінансові часові ряди часто демонструють ефекти типу ARCH.

Результати цих двох тестів виявляють присутність одиничного кореня у всіх рядах валютних курсів без винятку, що говорить про їх нестационарності. Треба відзначити, що цей висновок узгоджується з переважною більшістю досліджень, присвячених фінансовим часовим рядам. Другим висновком, який може бути зроблений, виходячи з проведених розрахунків, є стаціонарність цих рядів в перших різницях або відсутність одиничних коренів в рядах доходностей.

Тестування статистичних гіпотез має на увазі, що нульова гіпотеза відкидається тільки при наявності на це дуже серйозних підстав.

У разі, якщо дослідник хоче використовувати стаціонарність як нульової гіпотези і має підстави для переваги саме цього припущення, зручність традиційних тестів, перерахованих вище, виявляється під питанням. Квятковський, Філіпс, Шмідт, Шин (KPSS) розробили альтернативний підхід для перевірки наявності одиничних коренів, що базується на стаціонарності ряду в якості нульової гіпотези.

Також існує ще один тест, який належить до цієї групи і орієнтований безпосередньо на тестування наявності довгої пам'яті в якості альтернативної гіпотези - тест Ло (1991). Теверовські (1999) зазначив, що ймовірність прийняття нульової гіпотези про відсутність довгострокової залежності завжди виявляється переоціненою цим тестом, тим не менш, Кирман і Тейссієр (2002) показали, що цей тест є більш потужним, ніж тест KPSS.

Комбінування результатів, отриманих на основі статистик Діккея - Фуллера (і Філіпса - Перона) і KPSS (і Ло) призводить до чотирьох можливим наслідків:

- відкидання нульової гіпотези статистикою ADF і неможливість відкинути нульову гіпотезу статистикою KPSS розглядається як серйозне свідчення на користь стаціонарності ряду, тобто процесу $I(0)$;
- неможливість відкинути нульову гіпотезу тестом ADF і відкидання цієї гіпотези тестом KPSS служить важливим індикатором наявності одиничного кореня, тобто процесу $I(1)$;
- неможливість відкидання нульових гіпотез обома (ADF і KPSS) статистиками імовірно може свідчити про недостатній інформативності даних щодо довгострокових характеристик;
- одночасне відкидання нульових гіпотез статистиками ADF і KPSS вказує з високою часткою ймовірності на те, що даний процес не може бути описаний ні як процес $I(0)$, ні як $I(1)$.

Можна констатувати, що в нашому випадку ряди валютних курсів слідує, без сумніву, процесу $I(1)$, тоді як зробити остаточні висновки щодо рядів доходностей виявляється складніше. Тести ADF і PP однозначно відкидають гіпотезу про наявність поодиноких коренів у всіх рядах доходностей. Результати, отримані з використанням тесту KPSS, виявляються менш однозначними і чутливими до вибору модифікації цього тесту (з наявністю або відсутністю тренда). При цьому, як було вже сказано, тестові процедури надають перевагу нульовій гіпотезі, яка може бути відкинута тільки при серйозних свідцтвах не на її користь. У зв'язку з цим представляється можливим, що для процесів $I(d)$ з невеликими величинами параметра d , потужність тесту KPSS виявляється недостатньою для поділу процесів $I(0)$ і $I(d)$. При цьому тест Ло чітко вказує на відкидання гіпотези про короткої пам'яті для доходностей трьох

валютних курсів (французький франк, німецька марка і голландський гульден). Таким чином, можна говорити про високу ймовірність можливості опису трьох рядів доходностей процесом виду I (d). В роботі, однак, буде зроблена спроба оцінки моделей з довгою пам'яттю і для інших валют, оскільки, в силу обмеженості сили наявних тестів, свідчення, що підтримують припущення про коректність опису інших рядів процесом I (0), подаються не цілком достатніми.

3.1 Побудова моделі ARIFMA із потсійною дисперсією

В якості першого кроку процесу моделювання буде здійснено порівняння моделі "випадкового блукання" з низкою моделей з короткою і довгою пам'яттю за допомогою інформаційних критеріїв Акайке (AIC), Шварца (SC), Ханна - Куінна (HQIC).

Критерії різняться за величиною "штрафу", який накладається за включення додаткових параметрів (цей "штраф" мінімальний для критерію AIC і максимальний для SC).

Боллерслев і Міккелсен (1996) провели ряд симуляцій для порівняння AIC і SC, проте були змушені констатувати, що для різних процесів, що лежать в основі наявних даних, найкращий критерій вибору може змінюватися. Крім того, в якості допоміжного інструменту був використаний тест Боксу - Пірса, що дозволяє тестувати наявність кореляції в залишках моделі.

Результати оцінки моделей будуть приведені тільки для прибутковості французького франка. Табл. 3.2 показує, що модель ARFIMA виявляється найкращою в порівнянні з іншими альтернативними моделями за всіма використаними критеріями.

Таблиця 3.2 – Моделювання ARFIMA для FRF

	AIC	HQIC	SC
«Випадкове блукання»	2,05435	2,05530	2,05706
ARMA(1,0)	2,05339	2,05482	2,05746
ARMA(0,1)	2,05339	2,05482	2,05746
ARMA(1,1)	2,05381	2,05571	2,05923
ARMA(2,1)	2,05327	2,05565	2,06005
ARMA(1,2)	2,05335	2,05573	2,06012
ARMA(2,2)	2,05314	2,05500	2,06127
ARFIMA(0, ϵ , 0)	2,05297	2,05440	2,05703

Аналогічні висновки можуть бути зроблені і для курсів DEM-USD, NLG-USD і YEN-USD. Для остаточного вибору моделі був використаний тест правдоподібності (LRT), як було запропоновано Бейном, Лораном і Лекур (1999). Цей тест ґрунтується на порівнянні значень функцій правдоподібності для зіставляються моделей і розрахунку статистики $\varsigma = -2(L(\theta^0, y) - L(\theta^1, y))$ яка розподілена як χ^2 при нульовій гіпотезі про відсутність значущої різниці між значеннями функцій.

Необхідно відзначити, що такі порівняння мають сенс тільки в разі, коли одна модель зводиться до іншого (як, наприклад, модель ARFIMA (0, ξ , 0) і "випадкове блукання"). Кількість ступенів свободи розподілу дорівнює числу накладаються обмежень, необхідних для редукування однієї моделі до іншої.

Використання цього тесту показує, що для всіх трьох зазначених валютних курсів обмеження повинно бути відкинуто навіть при рівні значущості в 1%. Аналогічний аналіз для курсу GBP - USD і CAD - USD вказує на те, що тимчасові залежності тут також мають місце, однак вони носять короткостроковий

характер, і застосування інформаційних критеріїв і тесту LRT дозволяє нам вибрати для цих валют модель ARFIMA (1, 0, 0).

У разі курсу CHF - USD, модель "випадкового блукання" виявляється, без сумніву, найкращою і ознаки будь-яких залежностей повністю відсутні. Тест Боксу - Пірса показує, що для всіх валют обрані моделі дозволяють елімінувати автокореляцію залишків.

Оцінка моделей виду ARFIMA (p, ξ, q) для випадку, коли навіть найпростіша модель з довгою пам'яттю виявлялася переважно всіх інших альтернатив, які не будувалася, оскільки метою цього розділу був не вибір оптимальної моделі, а демонстрація того, що в певних випадках, навіть не враховуючи наявності часових залежностей в моментах більш високого порядку, моделі з довгою пам'яттю класу ARFIMA (а іноді і традиційні моделі з короткою пам'яттю) дозволяють описати поведінку рядів доходностей краще, ніж модель "випадкового блукання", подразумеваюч ася стандартної теорією.

Як було відзначено Тщерінгом (1994), з метою оцінки параметрів ARFIMA з використанням великої кількості спостережень облік структури GARCH не є необхідним.

Для випадку моделі AR - GARCH, Боллерслев (1986) показав, що оцінки параметрів авторегресії залишаються заможними і асимптотично ефективними, навіть коли структура GARCH не враховується. Тщерінг робить висновок про те, що цей результат залишається вірним і для більш загальних специфікацій типу ARFIMA в тому випадку, коли модель ARFIMA є оборотною і має нескінченну AR уявлення. Можна припустити, що розгляд моделі FIGARCH для дисперсії також не вплине на ці висновки.

Проте, необхідно сказати, що інші автори (Бейн, Лоран, Лекур (2002)), навпаки наполягають на важливості реалізації процедури вибору моделі тільки в класі повних моделей (зі змінною дисперсією), оскільки ними було виявлено

вплив умовної гетероскедастичності на вибір порядку AR і MA, так само як і на висновки щодо параметра часткової інтеграції.

У зв'язку з цим і з урахуванням того, що в наступному розділі буде запропоновано цілий ряд доповнень, спрямованих на поліпшення моделювання дисперсій, які потенційно можуть вплинути на вибір моделі для "середнього", будуть проведені нові порівняння моделей і перевірені гіпотези, що стосуються відсутності впливу на вихідні моделі для "середнього".

3.2 Побудова моделі ARIFMA – FIGARCH/FIAPARCH

Як уже було згадано раніше, багато дослідників відзначають, що волатильність фінансових часових рядів часто демонструє феномен кластеризації. Ця властивість може бути проілюстровано побудовою графіка квадратів помилок в часі (рис. 3.2).

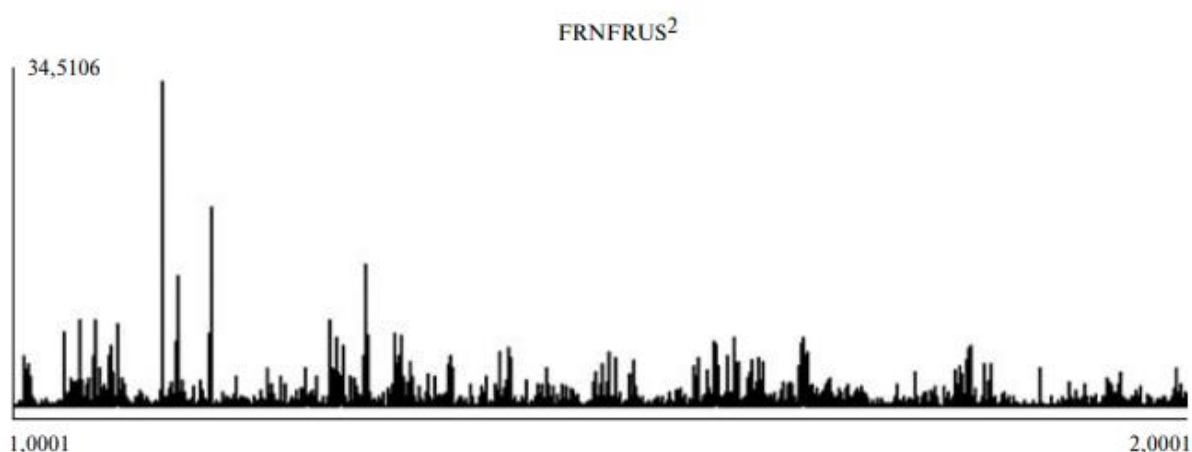


Рисунок 3.2 – Квадрати дохідностей обмінного курсу FRF – USD

Для перевірки наявності залежності волатильності від її попередніх значень (умовної гетероскедастичності) в ряду залишків моделі ARFIMA був використаний тест ARCH Енгл, заснований на тесті множників Лагранжа (LM) з нульовою гіпотезою про гомоскедастичність ряду. Для всіх оцінених моделей всіх, без винятку, валютних курсів, даний тест вказує на наявність структури типу ARCH в залишках.

Таким чином, представляє логічним використовувати моделювання типу ARCH для поліпшення якості отриманої моделі. Відповідно до ряду робіт останніх років, ми використовуємо модель з довгою пам'яттю в дисперсії - FIGARCH. Треба сказати, що традиційно використовуються моделі GARCH порядку не більше одиниці. Це стає тим більш важливим в разі моделі ARFIMA - FIGARCH, оскільки її оцінка займає тривалий час. Крім того, попередні дослідження з успіхом застосовували саме моделі FIGARCH низького порядку. Так, Бейлі та ін. (1996), досліджуючи курс DEM-USD і Тсе (1998), вивчаючи курс YEN-USD, вибрали в якості найкращої модель FIGARCH (1, ξ , 0), тоді як Бейн, Лоран, Лекур (2002) і Бейн, Лоран (2001) показали перевагу моделі FIGARCH (1, ξ , 1) для валютних курсів FRF-USD, DEM-USD, GBP-USD, YEN-USD. У даній роботі також будуть співставлені ці дві моделі для вибору оптимальної.

У той же час, як було показано в ході попереднього опису даних, що розглядаються, ряди виявляються розподіленими ненормально і, отже, необхідно брати до уваги також і цей факт. Для цієї мети буде зроблена спроба опису помилок моделей за допомогою трьох альтернативних розподілів розподіл Гаусса, розподіл Стюдента, асиметричний розподіл Стюдента). Як подальшого розширення моделі буде розглянута елестообразність включення в модель додаткових змінних. Треба відзначити, що у багатьох роботах (Палм, Влаар (1997), Бейлі, Боллерслев (1989)) включали дам-змінні, що представляють робочі дні тижня, а також святкові та вихідні дні як пояснюють змінних в моделях GARCH для денних валютних курсів і виявили, що включення цих "денних" ефектів дозволяло поліпшити специфікацію моделі. З цієї причини в

модель також були включені дві пояснюючі змінні, що представляють собою кількість святкових днів до і після поточного дня. Таке моделювання днів тижня має певну перевагу в порівнянні з традиційним моделюванням: це дозволяє сфокусуватися виключно на днях закриття і відкриття ринку, коли волатильність оходності очевидно може бути вище, при цьому не надто збільшуючи число оцінюваних параметрів. Порівняння різних моделей проводилося, як і раніше, за допомогою інформаційних критеріїв і тесту LRT. Докладні результати для кількох моделей представлені в табл. 3.3 для випадку французького франка.

Таблиця 3.3 – Альтернативні моделі для французького франка

Розподіл	$(0, d, 0)$	$(0, d, 0) - (1, \xi, 1)$			$(0, d, 0) - (1, \xi, 1)$			$(0, d, 0) - (1, \xi, 0)$			$(0, d, 0) - (1, \xi, 0)$		
		Гаусс	Стюдент		Гаусс	Стюдент		Гаусс	Стюдент		Гаусс	Стюдент	
		Без «ефекту дня»			3 «ефектом дня»			Без «ефекту дня»			3 «ефектом дня»		
μ	0,0074 (0,58)	0,0004 (0,04)	0,0144 (1,39)		0,0003 (0,03)	0,0122 (1,24)		0,0008 (0,07)	0,0144 (1,37)		0,0002 (0,02)	0,0116 (1,18)	
δ	0,0320 (2,86)	0,0299 (2,54)	0,0324 (3,04)		0,0235 (2,05)	0,0271 (2,59)		0,0321 (2,71)	0,0342 (3,22)		0,0233 (2,03)	0,0274 (2,62)	
ω	0,4556 (48,9)	0,0187 (4,16)	0,0176 (3,21)		-0,0288 (-4,80)	-0,0262 (-3,63)		0,0452 (6,32)	0,0419 (4,84)		-0,0027 (-0,34)	-0,0065 (-0,72)	
λ_1	—	—	—		0,0647 (5,80)	0,0548 (4,62)		—	—		0,0790 (6,63)	0,0700 (5,07)	
λ_2	—	—	—		0,0484 (5,27)	0,0534 (3,91)		—	—		0,0414 (4,51)	0,0505 (4,26)	
ξ	—	0,4345 (7,67)	0,4624 (6,45)		0,3936 (7,85)	0,4083 (6,54)		0,3230 (9,44)	0,3435 (7,49)		0,3005 (9,70)	0,3187 (8,02)	
β	—	0,5908 (10,33)	0,6117 (8,85)		0,5658 (9,83)	0,5629 (7,53)		0,2642 (6,93)	0,3011 (5,76)		0,2511 (7,29)	0,2803 (6,14)	
α	—	0,2215 (6,21)	0,2030 (4,64)		0,2292 (6,20)	0,2012 (4,20)		—	—		—	—	
ν	—	—	5,5413 (12,5)		—	5,7028 (12,3)		—	5,5308 (12,6)		—	5,6722 (12,45)	
AIC	2,0530	1,95257	1,88955		1,93661	1,87960		1,95647	1,89181		1,94043	1,8811	
HQIC	2,0544	1,95543	1,89288		1,94042	1,88389		1,95885	1,89467		1,94376	1,8849	
SC	2,0570	1,96071	1,89904		1,94746	1,89180		1,96325	1,89994		1,94992	1,8919	
Функція правдоподібності	-4896	-4653,8	-4502,4		-4613,7	-4476,7		-4664,1	-4508,8		-4623,8	-4481,2	

Аналіз табл. 3.3 дозволяє зробити ряд висновків. Найбільш важливим моментом є те, що оделирование дисперсії забезпечує дуже помітне поліпшення моделі в порівнянні з простою моделлю ARFIMA (в термінах значення функції правдоподібності й інформаційних критеріїв). Динаміка процесу для ряду доходностей досить слабо схильна до впливу включення різних форм залежностей для дисперсії (якщо розглядати абсолютні значення змін), проте це вплив існує.

Можна зробити висновок про те, що в разі, якщо дослідника цікавлять якісні висновки, то залежно в моментах порядку більшого одиниці, особливо для досить довгих рядів, можуть не враховуватися (що відповідає припущенням Тщерінга), однак для кількісних висновків і з метою прогнозування необхідно одночасно шукати оптимальну специфікацію рівняння для самого ряду і його дисперсій.

Це спостереження вірно також і для інших валютних курсів, оскільки довга пам'ять виявлена тільки в рядах, для яких цей феномен був присутній і в моделях з постійною дисперсією.

Використання розподілу Стюдента також веде до вкрай незначних змін коефіцієнтів, тоді як включення "денних ефектів" тягне за собою досить помітне зниження параметрів ξ і β . Останнє спостереження може бути пояснено тим фактом, що без включення денних ефектів ступінь персистентності виявляється штучно завищеною через те, що зростання дисперсії через цих ефектів (які є зумовленими, оскільки дати святкових днів відомі заздалегідь) розглядався як результат екзогенних шоків. Моделювання дисперсій процесом FIGARCH (1, ξ , 1) представляється кращим у порівнянні з процесом FIGARCH (1, ξ , 0) у всіх випадках.

Значення критеріїв AIC, HQIC, SC і функції правдоподібності були розраховані для цілого ряду альтернативних моделей з ефектами типу ARCH в дисперсіях і різними розподілами помилок для всіх валютних курсів. У табл. 3.4 представлені тільки остаточні моделі, обрані для розглянутих валютних курсів.

Таблиця 3.4 – Моделі, обрані для розглянутих валютних курсів

У дужках вказані t-статистики, * позначає значимість параметра на рівні 5%.

	Опис	FRF	DEM	NLG	CHF	YEN	CAD	GBP
μ	константа в рівнянні прибутковості	0,008 (0,35)	0,005 (0,47)	0,006 (0,56)	—0,003 (—0,36)	0,004 (0,25)	0,002 (0,57)	—0,009 (—1,34)
d	параметр довгої пам'яті в рівнянні прибутковості	0,027 (2,39)*	0,024 (2,27)*	0,024 (2,25)*	—	0,112 (3,43)*	—	—
Δ	"ефект січня" в рівнянні прибутковості	0,079 (2,37)*	0,086 (2,60)*	0,087 (2,65)*	0,104 (3,04)*	—	—	0,063 (2,41)*
ψ_1	параметр AR (1)	—	—	—	—	0,438 (6,16)*	0,043 (3,16)*	0,022 (1,62)
θ_1	параметр MA (1)	—	—	—	—	—0,565 (—7,12)*	—	—
ω	константа в рівнянні волатильності	0,011 (0,70)	—0,020 (—1,79)	—0,020 (—2,04)*	0,009 (0,74)	—0,016 (—1,82)	—0,003 (—2,40)*	—0,013 (—2,78)*
λ_1	ефект закриття ринку на волатильність	0,040 (3,57)*	0,056 (4,08)*	0,061 (3,70)	0,025 (1,94)	0,046 (3,42)*	0,002 (1,69)	0,024 (2,48)*
λ_2	ефект відкриття ринку на волатильність	0,050 (4,37)*	0,057 (4,83)*	0,050 (5,23)*	0,052 (4,16)*	0,05 (4,14)*	0,003 (2,16)*	0,029 (3,52)*
ξ	параметр довгою пам'яті в рівнянні волатильності	0,616 (6,59)*	0,651 (6,73)*	0,627 (5,73)*	0,533 (5,52)*	0,351 (7,63)	0,442 (6,12)*	0,442 (7,49)*
β	параметр "бета" моделі GARCH	0,717 (10,79)*	0,768 (13,89)*	0,750 (11,15)*	0,735 (12,51)*	0,594 (10,35)*	0,646 (11,39)	0,679 (14,90)*
α	параметр "альфа" моделі GARCH	0,148 (3,41)*	0,167 (3,18)*	0,172 (3,26)	0,246 (5,56)*	0,304 (6,51)*	0,299 (7,14)*	0,316 (8,31)*
γ	параметр асиметрії моделі APARCH	—0,012 (0,19)	—0,026 (—0,49)	—0,033 (—0,65)	0,028 (0,46)	—	—0,166 (—4,00)	—
δ	статичної параметр моделі APARCH	0,857 (4,77)*	1,325 (5,45)*	1,424 (6,22)*	1,194 (7,18)*	—	2,159 (12,75)*	—
ν	кількість ступенів свободи розподілу Стюдента	5,709 (12,23)*	6,775 (10,47)*	6,752 (10,40)*	6,581 (11,74)	4,681 (15,92)*	6,536 (12,78)*	6,047 (12,44)*
ς	коефіцієнт асиметрії в асиметричному розподілі Стюдента	—	—	—	—0,056 (—3,14)*	—0,076 (—4,25)*	—	—
Q(15)		21,24 (0,13)	19,59 (0,19)	19,91 (0,17)	12,31 (0,65)	26,23 (0,02)	17,79 (0,22)	29,48 (0,01)
Q2(15)		7,91 (0,85)	6,91 (0,91)	5,64 (0,96)	18,81 (0,13)	22,17 (0,05)	6,86 (0,91)	10,54 (0,65)

Параметр довгої пам'яті виявляється практично однаковим для трьох валют - FRF, DEM, NLG і рівним 0,027, 0,024 і 0,024 відповідно. Незважаючи на вкрай

низьку величину, він виявляється значущим у всіх трьох випадках на рівні значимості 5%. Незначність величини цього коефіцієнта може перешкодити здійсненню прогнозування, оскільки при такій величині параметра часткової інтеграції, прогнозні значення будуть мало відрізнятися від середнього значення ряду, проте, навіть в цьому випадку, факт виявлення довгою пам'яті зберігає своє значення, проливаючи світло на важливу закономірність поведінки фінансових часових рядів.

Параметр довгою пам'яті також виявився значущим, а його значення більш високим (0,112) для курсу YEN - USD, однак слід вкрай обережно інтерпретувати цей факт і стверджувати, що ступінь персистентності в даному випадку вище, оскільки для інших валютних курсів з довгою пам'яттю існує і друга досить конкурентоспроможна специфікація ARFIMA (1, d , 1) - FIGARCH (1, ξ , 1), при використанні якої значення параметра часткової інтеграції знаходяться в діапазоні від 0,09 до 0,1, що набагато ближче до значення, отриманого для японської ієни. Таким чином, представляється, що персистентність курсу YEN - USD все ж трохи вище, ніж для інших валютних курсів, але ступінь відмінності досить невелика.

Можна констатувати, що вибір кількості лагів для опису короткої пам'яті надає ключовий вплив на величину параметра довгою пам'яті і, швидше за все, більш коректним було б говорити про наявність феномена довгою пам'яті в цих чотирьох рядах і про достатню слабкості ступеня цієї пам'яті (персистентності) без зіставлення самих оцінених значень коефіцієнта.

Аналіз моделей для GBP, CAD і CHF вказує на відсутність слідів довгою пам'яті для цих валютних курсів. Моделі, вибрані для цих валют виявилися тими ж, що і для випадку з постійною дисперсією, тобто наявність короткої пам'яті підтверджується для перших двох валют, тоді як прибутковість курсу CHF абсолютно не залежить від попередніх значень.

Слід зазначити, що згідно з тестами на наявність поодиноких коренів в рядах доходностей валют FRF, NLG, DEM можна було припускати наявність довгої

пам'яті, тоді як щодо японської ієни такий висновок зробити не можна було. Така розбіжність може бути пояснено або удаваній довгою пам'яттю, яка характерна саме для використаної вибірки і не властива іншим періодам, або недостатньою силою використаних тестів (особливо в присутності одночасно довгої і короткої пам'яті). Так, наприклад, слабка сила тесту KPSS була виявлена Жірайтісом, Кокошка, Лейпусом і Тейссієром (2001) особливо в разі низької величини параметра часткової інтеграції, що і має місце в нашому випадку.

Таким чином, можна зробити висновок про персистентності прибутковості чотирьох валютних курсів (FRF, DEM, NLG, YEN), хоча її ступінь і виявляється вкрай невеликий. Висновки щодо наявності або відсутності довгої пам'яті в розглянутих валютних курсах, а також щодо величини цієї пам'яті в цілому відповідають результатам, отриманим Лекур (1999), Бейном, Лораном, Лекур (2002), Девідсоном (2002).

Незважаючи на значну кількість емпіричних робіт, існує значно менше досліджень, присвячених виробленню пояснень феномена довгою пам'яті і структурних моделей, які могли б генерувати поведінку такого роду. Проте, спробуємо запропонувати кілька альтернативних пояснень цього явища.

Зокрема, така поведінка може бути пов'язано з динамічними властивостями інших економічних змінних, таких як відносні ціни, відносна пропозицію грошей або відносні рівні доходу, які, як відзначається в багатьох дослідженнях, є частково інтегрованими.

У зв'язку з тим, що зміни в цих змінних лежать в основі коливань валютних курсів, що впливає з стандартних моделей формування валютного курсу, персистентність цих рядів може бути чинником, який сприяє присутності довгою пам'яті в рядах доходностей валютних курсів.

Ця гіпотеза має, однак, серйозний недолік: частота спостережень використовуваних макроекономічних рядів може відрізнятися від використовуваної частоти для валютного курсу (наприклад, місячні і денні дані).

Таким чином, встановлення суворої формальної взаємозв'язку між персистентного оходностей валютних курсів і макроекономічних змінних виявляється скрутним.

Якщо ж феномен довгої пам'яті дійсно міститься в самих лавах доходностей, а не в рядах фундаментальних змінних, що впливають на обмінний курс, то основним наслідком такого ефекту стає виникнення тривалого розриву між спостережуваним валютним курсом і його фундаментальним значенням. Цей феномен свідчить про тривалий період адаптації цін до нової інформації, що суперечить гіпотезі ефективного ринку. Можливе пояснення наявності довгої пам'яті на фінансових ринках укладено у взаємодії гетерогенних економічних агентів. Особливим випадком такої взаємодії можна вважати явище "наслідування" (mimic behavior), часто відзначається на фінансових ринках.

Дослідження, проведене Міньйон (1998), підтверджує і теоретичну можливість виникнення розриву між ціною фінансового активу і його фундаментальної вартістю, тобто "Бульбашки", який може виникати саме в силу наслідує "поведінки агентів. Міньйон показав, що якщо ступінь довіри економічних агентів до даних про стан фундаментальних змінних знижується, то оператори на ринку демонструють схильність до імітації поведінки інших. Як наслідок, шок, який породжує таку втрату довіри до фундаментальних змінним, матиме тривалі наслідки, тобто розрив між спостерігається ціною і фундаментальної вартістю активу буде зберігатися до тих пір, поки "міхур" не лопне або не «здується».

Таким чином, інтерпретація в термінах "наслідувального" поведінки може служити ще одним з можливих пояснень наявності довгої пам'яті на фінансових ринках. Феномен "бульбашок" виявляється, як правило, більш вираженим на фондових ринках, однак цей аналіз може бути поширений і на валютний ринок, оскільки психологія економічних агентів повинна залишатися аналогічної і, отже, "наслідувальне" поведінка може виникати і на цьому ринку. У світлі цієї гіпотези можна зробити цікавий коментар, що стосується наявності довгої пам'яті в рядах одних валютних курсів і її відсутності в інших. Свідомства присутності довгою

пам'яті не були виявлені для англійської та канадської валют, що можна пояснити наявністю тісних зв'язків між економіками цих країн і економікою США.

Ці валютні ринки виявляються, таким чином, більш прозорими в порівнянні з іншими, інформація на них поширюється швидше і невизначеність щодо фундаментальних змінних нижче. Подібне пояснення може бути запропоновано для випадку швейцарського франка, але замість тісних зв'язків між економіками, транспарентність цього ринку ґрунтується на меншій кількості його учасників, що працюють на ньому постійно і досить добре обізнаних про реалії ситуації.

Третє пояснення носить чисто технічний характер і пов'язане з агрегування або акумулюванням різних процесів надходження інформації. Ефект агрегування процесів надходження інформації також може генерувати персистентність доходностей валютних курсів.

Для поліпшення якості моделювання доходностей, в модель були включені Дамм-змінні, що дорівнюють 1 для спостережень, зроблених в січні і 0 для інших спостережень ("ефект січня" вже згадувався в першому розділі цієї роботи). Відповідна змінна виявилася значущою і дозволила поліпшити якість моделі для п'яти валют з семи. Значення змінної у всіх випадках позитивно, що вказує на збільшення прибутковості валютних курсів в січні в порівнянні з середньорічною прибутковістю. Оскільки валютний курс розглядається як кількість одиниць даної валюти за 1 долар США, то виявлений ефект означає, що американський долар, як правило, дорожчає в січні щодо інших валют, причому всі валюти, для яких даний ефект був виявлений виявляються європейськими (ефект відсутній для канадського долара і японської ієни).

Якщо для біржового ринку ми могли віднести цей феномен, наприклад, на рахунок надлишкового оптимізму ринкових операторів на початку кожного року, то знайти пояснення для випадку валютного ринку виявляється значно важче. Можна, можливо тільки сказати, що це пояснення може бути як об'єктивним, так і суб'єктивним. Під першим терміном мається на увазі можливість наявності

відмінностей в динаміці будь-яких макроекономічних змінних (ВВП, темп інфляції і т.д.), і в цьому випадку дамм-змінна грає роль проксі-змінної. Можливо також, що на початку кожного року багато інвесторів, в тому числі і інституційні, переглядають свої портфелі і здійснюють нові інвестиції більш активно, ніж протягом року.

При цьому багато європейських інвестори вважають за краще купувати американські активи, тоді як зворотний потік значно слабкіше, що створює надлишковий попит на американську валюту. Суб'єктивне пояснення може бути пов'язане з психологією економічних агентів, які з якої-небудь причини (наприклад, внаслідок більшої довіри до заяв представників американської влади) виявляються більш оптимістично налаштованими щодо перспектив американської економіки. Однак, даний феномен вимагає подальшого більш детального вивчення.

Результати розрахунків вказують на адекватний характер моделювання типу GARCH, оскільки воно дозволяє збільшити значення функції правдоподібності для всіх розглянутих моделей. Цей результат відповідає ряду попередніх досліджень, в яких успішно використовувалися моделі сімейства ARFIMA - GARCH. Проте, нам вдалося показати, що моделювання FIGARCH / FIAPARCH помітно перевершує моделювання GARCH. Цікаво, що Лекур (1999) отримала для ряду валютних курсів суму параметрів авторегресії і ковзного середнього моделі GARCH близьку до одиниці, що може мати на увазі наявність довгострокових часових залежностей в умовній волатильності. З метою перевірки того, чи робить вплив поточний шок на майбутню волатильність нескінченно довго, нею була оцінена модель IGARCH і зроблено висновок про її непридатність. У даній роботі також була здійснена оцінка моделі IGARCH і з використанням інформаційних критеріїв і тесту LRT, вона була відкинута. Параметр часткової інтеграції α є значущим на будь-якому розумному рівні значущості і його величина лежить між 0,35 і 0,65. Порядок цих цифр відповідає в цілому результатами, отриманими іншими авторами.

Кирман і Тейссієр (2002) запропонували мікроекономічна модель, що дозволяє пояснити феномен персистентності волатильності. Ця модель, як і в разі

персистентності самих доходностей, ґрунтується на наслідувальному поведінці економічних агентів. Автори описують можливий механізм виникнення такого роду поведінки, а також різких змін в думці операторів на ринку, які ведуть до виникнення "бульбашок". У той же час така модель генерує довгу пам'ять, і автори вказують на наявність тісного взаємозв'язку між довгою пам'яттю і "бульбашками" на фінансовому ринку.

Для дослідження того, чи мають позитивні і негативні шоки різний вплив на умовну дисперсію, була зроблена спроба застосування моделі FIAPARCH, яка виявилася вкрай успішною. Саме ця модель була обрана в якості найкращої специфікації для п'яти валют (крім YEN і GBP). Однак, основною перевагою даної моделі виявилася не її асиметрія, а гнучкість щодо вибору статистичного параметра для волатильності, який виявився, для більшості валют, відмінним як від двох, так і від одиниці, що узгоджується з результатами, отриманими Маккінзі і Мітчеллом (1998) .

При цьому єдиною валютою, для якої було виявлено асиметричний вплив позитивних і негативних шоків на динаміку волатильності, виявився канадський долар (аналогічне спостереження було зроблено Маккінзі і Мітчеллом (1998)). Пояснити подібне явище в разі валютних курсів непросто, однак можна припустити, що збільшення волатильності канадського долара при зниженні його вартості щодо долара США відбувається в результаті тиснотельної слабкості канадської економіки (в порівнянні з економікою США), що веде до виникнення побоювань учасників ринку щодо подальшого погіршення ситуації . Відсутності ефекту асиметрії для більшості валютних курсів і слід було очікувати, оскільки ця особливість поведінки відноситься переважно до інших видів фінансових активів (таким як, наприклад, акції).

Розрахунки підтверджують доцільність використання розподілу Стюдента, оскільки інформаційні критерії та тест правдоподібності вказують на помітну перевагу даного розподілу в порівнянні з нормальним. Величини t-статистик вказують на високу значимість параметра v . Також була зроблена спроба

використання асиметричного розподілу Стюдента, але для більшості валют (крім CHF і YEN) коефіцієнт асиметрії виявився незначну, а значення інформаційних критеріїв - гірше, ніж для аналогічних моделей, але з симетричним розподілом помилок.

Включення "ефекту дня" дозволило поліпшити якість моделі для всіх валют без винятку. Як і очікувалося, у зв'язку з тим, що відкриття торгів або очікування їх закриття може вплинути на ринкову активність, ці Дамм-змінні дійсно пояснюють деяку частину динаміки волатильності. При цьому представляється, що причини значущості двох використаних для опису "ефектів дня" змінних зовсім різні. Зростання волатильності при поновленні торгів після перерви пов'язаний, в першу чергу, з надходженням і адсорбцією нової інформації, що з'являється не тільки протягом поточного дня, а й накопиченої за час перерви в торгах (таким чином, інтенсивність цього процесу буде вищою, ніж в звичайні дні).

Цей результат є досить традиційним для аналізу волатильності фінансових ринків. Збільшення волатильності перед вихідними і святковими днями пояснити значно складніше. Констатація такого факту зустрічається набагато рідше, однак і це спостереження можна зустріти в недавніх роботах Лекур(1999), Бейна, Лорана, Лекур (2002) (на жаль, без будь-якого економічного коментаря). Можна запропонувати пояснювати цей феномен не факторами, безпосередньо пов'язаними з надходженням інформації, а психологією економічних агентів, які, знаючи заздалегідь, що негайна реаллокація ресурсів в період, коли ринок не функціонує, буде неможливою, починають проявляти занепокоєння, намагаючись вгадати можливі варіанти розвитку подій і зайняти оптимальну позицію перед обличчям очікуваної невизначеності.

У цих умовах поведінка цих агентів може відхилятися від раціонального. Введення дам-змінних для інших днів тижня не привело, згідно інформаційним критеріям, до поліпшення моделі. Щоб уникнути накладення апріорних обмежень, ті ж дам-змінні були використані і для рівняння умовного середнього, проте всі вони виявилися незначними.

Як вже було зазначено, інформаційні критерії і значимість коефіцієнтів для всіх валют в моделі FIGARCH / FIAPARCH дозволили зробити вибір на користь опису процесу динаміки волатильності за допомогою модифікацій FIGARCH (1, ξ , 1) і FIAPARCH (1, ξ , 1) з "денними ефектами" і ненормальним розподілом помилок (Стюдента і асиметричного Стюдента для різних валют). Важливе спостереження, яке може бути зроблено на основі отриманих результатів, полягає в вкрай схожих оцінках моделей для обмінних курсів валют трьох країн, які увійшли в зону євро. Це дозволяє припустити, що сама єдина європейська валюта може вести себе аналогічним чином. Перевірка цієї гіпотези є досить цікавою, проте вона стане можливою тільки після накопичення ряду обмінних курсів EUR - USD достатньої довжини.

Для того щоб модель могла бути використана для прогнозування або для опису механізму, що лежить в основі зміни цін, необхідно, щоб оцінені параметри зберігали стабільність на протязі певного періоду. Тестування стабільності було проведено з використанням тесту Ніблома, запропонованого Лі і Хансеном (1992). Статистика Ніблома тестує нульову гіпотезу про сталість вектора коефіцієнтів в цілому при альтернативній гіпотезі про його нестабільності і має розподіл, залежне тільки від числа параметрів в векторі коефіцієнтів. Крім того, існує і статистика, що тестує індивідуальні параметри. Результати тесту вказують на разючу стабільність коефіцієнтів для всіх валют, оскільки нульова гіпотеза не може бути відкинута на рівні 1%, а в багатьох випадках і 5%. Така стабільність є ще одним аргументом на користь вірного вибору специфікації моделей.

Тест автокореляції Боксу - Пірса, застосований до самих рядів доходностей і їх квадратах, вказує на відсутність залишкової автокореляції помилок, а результати тесту ARCH свідчить про те, що наявна умовна гетероскедастичності врахована в моделі в повному обсязі.

Методи, використані в цьому розділі, дозволили продемонструвати цілий ряд "хороших" властивостей побудованих моделей, проте досить часто якість "підгонки" в рамках вибірки переоцінює прогнозну силу моделі в результаті

надмірного наближення до наявних даним (over-adjustment). З цієї причини питання побудови прогнозів виявляється вкрай важливим і повинен бути досліджений окремо.

3.3 Побудова прогнозу дохідності

У попередньому розділі було показано, що процеси класу ARFIMA дійсно описують частина поведінки доходностей деяких з валютних курсів. У цьому розділі буде зроблено спробу з'ясувати, чи можуть ці прибутковості бути передбачені на основі таких процесів, тобто виходячи з попередніх значень доходностей. Як було відзначено вище, "хороші" властивості моделі в межах вибірки не завжди ведуть до отримання аналогічних характеристик для прогнозів, зроблених за межами вибірки. Аналіз прогнозної сили моделі буде проведено шляхом зіставлення якості прогнозів, одержуваних на основі обраних в попередньому розділі моделей і "наївних" прогнозів, отриманих з моделі "випадкового блукання" для валютного курсу.

Таким чином, як і в деяких інших авторів, була зроблена спроба перевершити "випадкове блукання" з точки зору якості одержуваних прогнозів. Припущення про те, що курс слідує процесу "випадкового блукання" означає, що "найкращим" прогнозом валютного курсу на момент $(t + 1)$ є курс, що спостерігався в момент t . У разі якщо таке припущення справджується, моделювання валютного курсу з метою прогнозування на основі досить складних процесів втрачає сенс, оскільки лише останній спостерігався валютний курс містить інформацію, придатну для прогнозування курсу в наступний момент часу.

Метою цього розділу і є з'ясування того, чи може прийняття до уваги ефекту довгої пам'яті привести до помітного поліпшення результатів прогнозування. При цьому треба пам'ятати про те, що оцінка параметра d залежить від помилок апроксимації і схильна до деякого зсуву, пов'язаного з "урізанням" полінома $(1 - L)^d$

. Це недосконалість використовуваних для оцінки методик може призвести до деякого погіршення якості прогнозів, одержуваних з використанням моделі ARFIMA.

Для цілей прогнозування були використані валютні курси, для яких параметр часткового інтегрування виявився значним за результатами оцінки моделей ARFIMA - FIGARCH / FIAPARCH. З метою розрахунку прогнозних значень за межами вибірки моделі були переоцінені кілька разів - для вибірок, що закінчуються груднем 1996 і груднем 1997 року для курсів FRF, DEM, NLG і закінчуються груднем 1999 року, груднем 2000 року і серпні 2002 р для YEN. На основі цих оцінок будувалися прогнози на 1997 і 1998 рр. для перших трьох валют і на 2000, 2001 рр. і перше півріччя 2002 р - для ієни. Результати проведеного переоцінювання виявилися практично ідентичними оцінками, отриманими в попередньому розділі з використанням повної вибірки, що підтверджує висновки, зроблені на основі тесту Ніблома. Для порівняння прогнозів, отриманих за моделлю ARFIMA - FIGARCH / FIAPARCH з прогнозами, заснованими на моделі "випадкового блукання" була використана величина середньоквадратичної помилки (RMSE):

Величина цього показника для обраних в попередньому розділі моделей і для

$$RMSE = \left[\sum_{i=0}^{T_k-1} \frac{1}{T_k} (\tilde{X}'_{t+i+k} - X_{t+i+k})^2 \right]^{\frac{1}{2}}$$

моделі "випадкового блукання" для різних прогнозних періодів приведена в табл. 3.5.

Таблиця 3.5 – Порівняння прогнозної сили моделей «випадкового блукання» та ARFIMA – FIGARCH/FIAPARCH

Обмінний курс	Період		RMSE
FRF – USD	1997	“Випадкове блукання”	0,6000
		ARFIMA – FIGARCH	0,5981

	1998	“Випадкове блукання” ARFIMA – FIGARCH	0,5443 0,5440
DEM – USD	1997	“Випадкове блукання” ARFIMA – FIGARCH	0,6111 0,6094
	1998	“Випадкове блукання” ARFIMA – FIGARCH	0,5455 0,5447
NLG – USD	1997	“Випадкове блукання” ARFIMA – FIGARCH	0,6228 0,6210
	1998	“Випадкове блукання” ARFIMA – FIGARCH	0,5402 0,5393
YEN – USD	2000	“Випадкове блукання” ARFIMA – FIGARCH	0,6327 0,6312
	2001	“Випадкове блукання” ARFIMA – FIGARCH	0,6106 0,6093
	2002	“Випадкове блукання” ARFIMA – FIGARCH	0,6564 0,6559

Дані табл. 3.5 дозволяють зробити висновок про те, що в цілому розроблені вище процеси з довгою пам'яттю володіють прогностою силою, оскільки вони виявляються в стані перевершити випадкове блукання згідно із запропонованими критеріями. Різниця між двома моделями в термінах помилки прогнозу є вкрай слабким, проте, незважаючи на це, важливо відзначити систематичний характер переваги процесу ARFIMA - FIGARCH / FIAPARCH. Створюється враження, що отриманий результат має безпосередній вплив на дискусію про спроможності гіпотези ефективного ринку, відповідно до якої майбутні прибутковості не можуть бути передбачені на основі їх попередніх значень. Оскільки використані моделі з довгою пам'яттю дозволяють отримати прогнози більш високої якості, ніж "наївні" прогнози, складаються в збереженні в ролі очікуваної в наступному періоді прибутковості її попереднього значення, що спостерігалось на ринку, гіпотеза

ринкової ефективності виявляється поставленої під сумнів. Таке твердження відповідає результатам, отриманим Бейном, Лораном, Лекур (2002), Лардік, Мін'йон (1999), Лекур (1999), Мін'йон (1998), Тщерінгом (1994) і деякими іншими авторами. При цьому необхідно підкреслити, що зазначене поліпшення якості прогнозів носить виключно статистичний характер і в разі, якщо ми звернемося до подання про ефективний ринок, запропонованим в першому розділі, відповідно до якого механізм корекції виникаючих неефективно починає функціонувати тільки коли можливо їх використання для отримання додаткового прибутку, то після врахування вартості виконуваних операцій трансакційних витрат слабка перевага процесів з довгою пам'яттю може виявитися повністю елімінувати і, в цьому трактуванні, гіпотеза ефективного ринку зберігає своє значення.

Висновок до розділу

В даному розділі було розглянуто поведінку доходностей курсів валют ряду країн щодо долара США. В ході дослідження була піддана перевірці гіпотеза про можливу наявність у даних рядах довгої пам'яті, будували застосовані на практиці моделі і методи, описані в попередній частині роботи, з метою вибору моделей, найбільш повно і адекватно описуючих реалії сучасних валютних ринків різних країн.

РОЗДІЛ 4 РОЗРОБКА СТАРТАП-ПРОЕКТУ

Стартап – це тимчасова організація, створена для пошуку повторюваної, масштабованої і стійкої бізнес-моделі. Стартап не є маленькою копією великої компанії. Це бізнес, який будується для подальшого продажу і ґрунтується на дещо інших правилах. Цим він і відрізняється від традиційного бізнесу. Стартап спочатку передбачає наявність якоїсь інновації, наприклад, заснованої на новій технології, або інновації в бізнес-процесах, або адаптації рішення, якого до цього не було на локальному ринку. Це означає, що стартапер перебуває в стані високої невизначеності. Тобто існують припущення щодо майбутніх клієнтів і партнерів, постачальників, зв'язків тощо. Відповідно, інструментарій і кроки, які потрібно буде зробити, дещо відрізняються від традиційного бізнесу.

Після того як буде знайдено повторювану бізнес-модель, бізнес потрібно масштабувати. Ось тут і знадобляться інвестиції, які стануть необхідним паливом для стрімкого зростання. Адже пройти шлях від компанії в гаражі до компанії вартістю кілька мільйонів або мільярдів доларів без зовнішніх інвестицій у більшості випадків майже неможливо. Розроблення та виведення стартап-проекту на ринок передбачає здійснення чотирьох етапів: маркетинговий аналіз, організація, фінансово-економічний аналіз та заходи з комерціалізації стартап-проекту.

Даний розділ магістерської дисертації присвячено реалізації першого етапу розробки стартап-проекту, а саме висвітленню маркетингових аспектів створення стартапу: відбору ідей, створенню концепції продукту, визначення перспектив ринкової реалізації проекту та розроблення маркетингової стратегії.

Метою розділу є проведення маркетингового аналізу стартап проекту задля визначення принципової можливості його ринкового впровадження та можливих напрямів реалізації цього впровадження.

4.1 Опис ідеї проекту

Даний стартап-проект стосується розробки та реалізації програмного забезпечення для моделювання та прогнозування процесів із довгою пам'яттю. Системи з використанням цих алгоритмів для зазначеної задачі раніше розроблені не були. Реалізація алгоритмів буде проводитись на мові програмування, призначеній саме для статистичного аналізу даних R у середовищі RStudio. Переваги та вигоди від продукту наведено у таблиці 4.1.

Таблиця 4.1 – Опис ідеї стартап-проекту

Зміст ідеї	Напрямки застосування	Вигоди для користувача
Розробка ПО для моделювання та прогнозування процесів із довгою пам'яттю	1. Фінансові установи: банки, страхові компанії, біржі	Можливість побудови моделей та подальшого прогнозування процесів із довгою пам'яттю
	2. Гідрологічні, геологічні та екологічні дослідження	

Для аналізу відмінностей даного продукту від аналогів було визначено перелік техніко-економічних властивостей та характеристик ідеї, а також означено попереднє коло конкурентів.

На даний момент продукт не має серйозних конкурентів на вітчизняному ринку. Зарубіжні аналітико-інформаційні системи не можуть бути придатними у вітчизняних реаліях. Результати порівняльного аналізу стартап-ідеї та конкурентів наведено у таблиці 4.2.

Таблиця 4.2 – Визначення сильних, слабких та нейтральних характеристик ідеї проекту

№	Техніко-економічні характеристики ідеї	Потенційні конкуренти	Мій проект	Слабка сторона	Нейтральна сторона	Сильна сторона
1	Ціна додатків, створених за допомогою платформи	100 000	20 000 – 40 000	Потрібні великі обсяги даних		Низька ціна
2	Час створення додатку за допомогою платформи	3 місяці	3 місяці			Швидкість розробки

Продовження таблиці 4.2.

3	Сумісність з іншими ОС	Відповідно вимогам замовника	Сумісність з Windows		Можливість створення для різних операційних систем	
4	Відповідність ситуації на державному ринку	Не відповідає	Відповідає			Алгоритми будуть врахувати реакцію на різні події саме української аудиторії

4.2 Технологічний аудит ідеї проекту

Маркетинговий аналіз стартап-проекту передбачає проведення аудиту технології, за допомогою якої можна реалізувати ідею проекту. Результати аналізу технологічної здійсненності проекту наведено у таблиці 4.3.

Таблиця 4.3 - Технологічна здійсненність ідеї проекту

№	Ідея проекту	Технологія реалізація	Наявність технології	Доступність
1	Розробка ПО для моделювання та прогнозування процесів із довгою пам'яттю	Мова програмування – R, RStudio - середовище розробки	Наявна	Доступна

За результатами таблиці 3 можна зробити наступні висновки: для реалізації головної ідеї стартап-проекту наявне на ринку та доступне середовище програмування RStudio, яке дозволяє легко реалізувати моделі, необхідні для досягнення мети та ідеї проекту та є безкоштовним. Тому для реалізації ідеї стартапу було використано середовище розробки RStudio.

4.3 Аналіз ринкових можливостей запуску стартап-проекту

Щоб спланувати напрями розвитку проекту із урахуванням стану ринкового середовища, необхідно визначити ринкові можливості, які можна використати під час ринкового впровадження проекту та ринкові загрози, що можуть перешкодити реалізації проекту.

Спочатку проводиться аналіз попиту. Результати вказані у таблиці 4.4.

Таблиця 4.4 – Попередня характеристика потенційного ринку стартап-проекту

№	Показники стану ринку (найменування)	Характеристика
1	Кількість головних гравців, од	3
2	Загальний обсяг продаж, грн/ум.од	6 500 000 у.о.
3	Динаміка ринку (якісна оцінка)	Зростає
4	Наявність обмежень для входу (вказати характер обмежень)	Немає
5	Специфічні вимоги до стандартизації та сертифікації	Немає
6	Середня норма рентабельності в галузі (або по ринку), %	15

За результатами таблиці 4 можна зробити висновок про те, що ринок є привабливим для входження.

Надалі визначаються потенційні групи клієнтів, їх характеристики, та формується орієнтовний перелік вимог до товару для кожної групи (табл. 4.5).

Таблиця 4.5 - Характеристика потенційних клієнтів стартап-проекту

№	Потреба, що формує ринок	Цільова аудиторія	Відмінності у поведінці різних потенційних цільових груп	Вимоги споживачів до товару
1	Аналіз і прогнозування екологічного стану територій	Організації захисту навколишнього середовища	Розмір бюджету організації. Наявність можливості співробітництва з компанією, що розробляє стартап	Захищеність даних, точність обчислень та прогнозу
2	Прогнозування поведінки економічного ринку	Фінансові установи: банки, страхові компанії, біржі	Розмір бюджету організації. Наявність можливості співробітництва з компанією, що розробляє стартап	Захищеність даних, точність обчислень та прогнозу

Після визначення потенційних груп клієнтів проводиться аналіз ринкового середовища: складаються таблиці факторів, що сприяють ринковому впровадженню проекту, та факторів, що йому перешкоджають (таблиці 4.6 – 4.8).

Таблиця 4.6 – Фактори загроз

№ п/п	Фактор	Зміст загрози	Можлива реакція компанії
1	Війна	Відсутність попиту в умовах ведення бойових дій	Зміна вектору збуту товару
2	Ресурси	Відсутність даних для аналізу	Пошук організацій, які можуть надати дані
3	Проблеми зі збутом	Висока ціна	Відстежувати рівень цін конкурентів
4	Політична криза	Рівень корумпованості влади	Пошук клієнтів на міжнародному ринку

Таблиця 4.7 – Фактори можливостей

№ п/п	Фактор	Зміст можливості	Можлива реакція компанії
1	Залучення нових джерел інформації	Поява часткового доступу до великої кількості даних зі злочинності	Більш детальне дослідження даних
2	Збут товару	Передбачається проведення тендеру на закупівлю аналітичного ПЗ	Участь у тендері

Таблиця 4.8 – Ступеневий аналіз конкуренції на ринку

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
1. Тип конкуренції - чиста	Не існує монополії на розробку аналітичних продуктів	Сприяє вільній розробці та реалізації продукту
2. За рівнем конкурентної боротьби міжнаціональний	Продукт повинен показати кращі результати, ніж закордонні аналоги	Ведучи конкуренцію на національному рівні, необхідно прикласти належні зусилля для охоплення всього національного ринку.
3. За галузевою ознакою внутрішньогалузева	Конкуренція стосується ринку аналітичних продуктів	Необхідно ретельно врахувати особливості бізнесу, для якого розроблюється система, щоб бути конкурентоспроможним
4. Конкуренція за видами товарів: - товарно-родова	Конкуренція щодо задоволення потреб зручної обробки даних з відвідування матчів.	Конкурентна боротьба з широковідомими продуктами для аналізу даних(Excel, SPSS).

Продовження таблиці 4.8

5. За характером конкурентних переваг нецінова	Є безкоштовні системи для аналізу даних, але вони не є спеціалізованими саме для потреб потенційних споживачів	Необхідно зробити системи зрозумілою для користування та врахувати специфічні потреби
6. За інтенсивністю - марочна/не марочна	Не марочна	Не існує аналогів системи відомої марки

Далі проводиться детальний аналіз умов конкуренції в галузі (таблиця. 4.9), що дозволяє визначити найкращу відповідність між внутрішнім станом організації і дією сил у її зовнішньому оточенні.

Суперництво між існуючими фірмами. Конкуренція між фірмами, що суперничають і які пропонують однотипні товари і послуги, виникає у зв'язку з тим, що в одній чи декількох фірм з'являється можливість краще задовольнити потреби споживача або необхідність поліпшити свою діяльність. До основних засобів конкурентної боротьби можна віднести: більш низькі ціни; поліпшені характеристики товару; більш високий рівень обслуговування споживачів; тривалі терміни гарантійного періоду; спеціальні способи просування товару на ринок; випуск нових товарів; використання слабкостей конкурентів.

Конкурентна сила покупців зростає з підвищенням їх можливості впливати на ціни і якість товарів, на рівень обслуговування та ін.

Застосування для аналізу конкуренції в галузі моделі п'яти конкурентних сил М. Портера дозволяє визначити структуру цих сил, оцінити кожен силу і приступити до формування конкурентної стратегії.

Таблиця 4.9 – Аналіз конкуренції за Портером

	Потенційні конкуренти	Клієнти	Товари-замінники
Складові аналізу	Вітчизняні компанії, що займаються розробкою ПЗ	Клієнти мають досить великий бюджет	Можливість випуску специфічного безкоштовного ПЗ, що задовільнить потреби клієнта
Висновки	Є можливість виходу на ринок та потенційні конкуренти	Клієнти повністю диктують умови роботи, бо від їх потреби та побажання формують попит на розробку специфічного ПЗ	Обмеження є з боку пакетів для широкого аналізу даних, але вони не значні, бо не вирішують необхідних клієнтам завдань

З урахуванням характеристик ідеї проекту (таблиця 4.2), вимог споживачів до товару (табл. 4.5) та факторів маркетингового середовища (таблиці 4.6- 4.7) визначається та обґрунтовується перелік факторів конкурентоспроможності. Конкурентоспроможність - ринкова категорія, що має динамічний, мінливий характер. Так, при незмінних якісних характеристиках товару, його конкурентоспроможність може змінюватися в широких межах залежно від кон'юнктури ринку, дій конкурентів, зміни цін, рекламних заходів тощо.

Конкурентоспроможність підприємства повинна забезпечуватися такими показниками, як висока виробнича ефективність, яка реалізується завдяки сучасному обладнанню, технологіям, кваліфікованими працівниками та здатністю

завоювати й тривалий час утримувати стійні позиції на ринку, що забезпечується завдяки ефективному використанню принципів маркетингового управління.

Аналіз наводиться в таблиці 4.10.

Таблиця 4.10 – Обґрунтування факторів конкурентоспроможності

№ п/п	Фактор конкурентоспроможності	Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим)
1	Низька собівартість продукції	Реалізація алгоритму порівняно невисока й мало затратна за часом й трудовими ресурсами
2	Висока точність прогнозу	Результати роботи алгоритму показують високі значення прогнозних статистик.
3	Широкий спектр застосування	Алгоритм можна застосовувати для аналізу й прогнозування багатьох процесів із довгою пам'яттю

За визначеними факторами конкурентоспроможності (таблиця 4.10) проводиться аналіз сильних та слабких сторін стартап-проекту (таблиця 4.11).

Таблиця 4.11 – Порівняльний аналіз сильних та слабких сторін «назва проекту»

№ п/п	Фактор конкурентоспроможності	Бали 1-20	Рейтинг товарів-конкурентів у порівнянні з SASSimilarTrajection						
			-3	-2	-1	0	+1	+2	+3
1	Низька собівартість продукції	12							x
2	Висока точність прогнозу	17	x						
3	Широкий спектр застосування	18				x			

Фінальним етапом ринкового аналізу можливостей впровадження проекту є складання SWOT-аналізу (матриці аналізу сильних (Strength) та слабких (Weak) сторін, загроз (Troubles) та можливостей (Opportunities) (таблиця 4.12) на основі виділених ринкових загроз та можливостей, та сильних і слабких сторін (таблиця 4.11).

Таблиця 4.12 –SWOT аналіз стартап-проекту

<p>Сильні сторони:</p> <ul style="list-style-type: none"> - низька собівартість розробки - висока точність роботи алгоритму - широкий спектр застосування 	<p>Слабкі сторони:</p> <ul style="list-style-type: none"> - необхідність мати вірні дані
<p>Можливості:</p> <ul style="list-style-type: none"> - залучення нових джерел інформації - вихід на ринок інших фінансових установ 	<p>Загрози:</p> <ul style="list-style-type: none"> - війна - ресурси - політична криза

На основі SWOT-аналізу розробляються альтернативи ринкової поведінки для виведення стартап-проекту на ринок та орієнтовний оптимальний час їх ринкової реалізації.

Визначені альтернативи аналізуються з точки зору строків та ймовірності отримання ресурсів (таблиця 4.13).

Таблиця 4.13 – Альтернативи ринкового впровадження стартап-проекту

№ п/п	Альтернатива (орієнтовний комплекс заходів) ринкової поведінки	Ймовірність отримання ресурсів	Строки реалізації
1	Розробка додаткових модулів, що дозволить використовувати систему не тільки в контексті надання кредитів	Середня	1 рік

4.4 Розроблення ринкової стратегії проекту

Стратегія маркетингу – це комплекс базових рішень, спрямованих на досягнення генеральної мети фірми і вихідних з оцінки ринкової ситуації і власних можливостей, а також інших факторів та сил навколишнього середовища маркетингу. Маркетингова стратегія компанії на сучасному ринку формується під впливом безлічі факторів. Реалізація стратегії сприяє перетворенню продукту в товар. Розроблення ринкової стратегії першим кроком передбачає визначення

стратегії охоплення ринку: опис цільових груп потенційних споживачів (таблиця 4.14).

Таблиця 4.14 –Вибір цільових груп потенційних споживачів

№ п/п	Опис профілю цільової групи потенційних клієнтів	Готовність споживачів сприйняти продукт	Орієнтовний попит в межах цільової групи (сегменту)	Інтенсивність конкуренції в сегменті	Простота входу у сегмент
1.	Фінансові установи: банки, страхові компанії	Середня	Високий	Середня	Складна
2.	2. Гідрологічні, геологічні та екологічні дослідження	Висока	Середня	Низька	Низька

За результатами аналізу потенційних груп споживачів (сегментів) доцільно вибрати конкретну цільову групу та відповідно до неї визначити стратегію охоплення ринку. Оскільки, цільовою аудиторією на яку розрахований продукт, можуть бути організації, зацікавлені в отриманні максимального прибутку від проведення матчу є вибір стратегії цільового маркетингу.

Вибір стратегії-центральний момент стратегічного планування. Кульмінаційним моментом вибору стратегії є аналіз і оцінка альтернативних варіантів. Завдання оцінки полягає у розробці такої стратегії, що забезпечувала б максимальну ефективність роботи в майбутньому.

Стратегічний вибір повинен бути заснований на чіткій концепції розвитку проекту, а саме: формулювання - однозначним і ясным, тому що обрана стратегія

на тривалий час обмежує волю дій керівництва і впливає на всі прийняті ним рішення. Тому обрана альтернатива ретельно досліджується й оцінюється. При цьому мають братися до уваги численні фактори: ризик, досвід минулих стратегій, вплив власників акцій, фактор часу тощо.

Для роботи в обраних сегментах ринку необхідно сформувати базову стратегію розвитку (таблиця 4.15).

Таблиця 4.15 –Визначення базової стратегії розвитку

№ п/п	Обрана альтернатива розвитку проекту	Стратегія охоплення ринку	Ключові конкурентоспромож ні позиції відповідно до обраної альтернативи	Базова стратегія розвитку*
1.	Розширення цільової аудиторії	Розробка додаткових модулів, що дозволить використовувати методологію в інших напрямах	Врахування особливостей	Стратегія диференціації

Наступним кроком є вибір стратегії конкурентної поведінки (таблиця 4.16). Розробляючи конкурентну стратегію, керівники прагнуть знайти і втілити спосіб вигідно і довготривало конкурувати в свій галузі. Універсальної конкурентної стратегії не існує; лише стратегія, що узгоджена з умовами конкретної галузі, досвідом та бюджетом, яким володіє конкретний проект, може принести успіх.

Таблиця 4.16– Визначення базової стратегії конкурентної поведінки

№ п/п	Чи є проект «першопрохідцем» на ринку?	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристики товару конкурента, і які?	Стратегія конкурентної поведінки*
1	Проект вдосконалює уже існуючі методи аналізу й прогнозування, враховує специфіку галузі	Залучення нових споживачів шляхом розширення діяльності компанії на різні ринки збуту	- якість прогнозування - робота з різномісними даними	Стратегія виклику лідера

На основі вимог споживачів з обраних сегментів до постачальника (стартап-компанії) та до продукту (див. таблицю 4.5), а також в залежності від обраної базової стратегії розвитку (таблиця 4.15) та стратегії конкурентної поведінки (табл. 4.16) розробляється стратегія позиціонування (таблиця 4.17). що полягає у формуванні ринкової позиції (комплексу асоціацій), за яким споживачі мають ідентифікувати торгівельну марку/проект.

Таблиця 4.17 –Визначення стратегії позиціонування

№ п/п	Вимоги до товару цільової аудиторії	Базова стратегія розвитку	Ключові конкурентоспроможні позиції власного стартап-проекту	Вибір асоціацій, які мають сформувати комплексну позицію власного проекту (три ключових)
1	Висока точність результату	Окремо досліджувати кожну галузь	Точність буде вища за ту, що можуть забезпечити інші системи без врахування специфіки галузі	Точність, зручність, гнучкість

Позиціонування залежить від вибору цільового ринку та створення комплексної пропозиції для залучення й задоволення цього (цільового) ринку краще, ніж це роблять конкуренти. Після визначення поточної позиції та спрямованості її розвитку розглядаються кілька основних стратегічних альтернатив.

Висновок до розділу 4

Розділ мав на меті проведення маркетингового аналізу стартап-проекту задля визначення принципової можливості його ринкового впровадження та можливих напрямів реалізації цього впровадження.

Проведений аналіз має визначальне значення для оцінки життєздатності проекту, тому що дозволяє одержати стратегічно важливу ринкову інформацію

ВИСНОВОК ПО РОБОТІ

В даному дослідженні зроблена спроба досягнення двох тісно взаємопов'язаних цілей. З одного боку, були зіставлені різні моделі часових рядів для виявлення їх порівняльних характеристик, вибору найкращої і тестування передбачуваності майбутніх значень фінансових змінних. З іншого боку, була розглянута гіпотеза ефективного ринку і зроблено спробу її перевірки. Для досягнення цих цілей використаний цілий ряд модифікацій моделей з подвійною довгою пам'яттю (включаючи модель ARFIMA - FIAPARCH, яка вперше введена в даній роботі) з трьома різними розподілами залишків (асиметричний розподіл Стюдента також було вперше застосовано для випадку моделей з довгою пам'яттю). Для чотирьох валютних курсів з семи були виявлені свідчення присутності довгою пам'яті в рядах прибутковості, хоча ступінь персистентності і виявилася вкрай слабкою. При виборі оптимальної моделі в класі ARFIMA - FIGARCH / FIAPARCH була виявлена необхідність використання кілька відмінних специфікацій як для рівняння доходностей, так і для рівняння, що описує динаміку волатильності. Для всіх валют (для яких була виявлена довга пам'ять), за винятком японської ієни, нова модель ARFIMA - FIAPARCH виявилася найкращою, незважаючи на той факт, що параметр асиметрії виявився в більшості випадків незначним (симетричне вплив шоків на волатильність валютних курсів збігається з результатами ряду попередніх досліджень). Помітне поліпшення моделей для всіх валют було досягнуто за рахунок застосування розподілів з "товстими хвостами", тоді як помітна асиметрія розподілу інновацій присутній тільки в разі японської ієни і швейцарського франка.

Включення ефектів відкриття і закриття ринку дозволило зробити висновок про помітне зростання волатильності в ці періоди. В роботі запропоновані можливі пояснення більшості зазначених в ході проведеного аналізу феноменів і з використанням статистичних тестів виявлено цілу низку властивостей моделі, що

дозволяють розглядати характер здійсненого моделювання як адекватний. При цьому тести, проведені поза вибіркою, демонструють помітно менш задовільні результати. Внаслідок малої величини параметра часткового інтегрування в рівнянні для доходностей, прогнози, отримані на базі побудованої моделі, виявляються лише трохи краще "наївних" прогнозів на основі моделі "випадкового блукання". На відміну від висновків деяких інших дослідників, автор припускає, що цей факт не обов'язково суперечить гіпотезі ефективного ринку, оскільки вважається за необхідне враховувати економічний, а не статистичний вииграш від такої прогнозованості. В роботі запропоновано розуміння ефективного ринку як механізму, що коректують завдяки діям учасників ринку, які починають діяти тільки при наявності можливості отримання додаткового прибутку, у зв'язку з чим певна ступінь прогнозованості теоретично може бути присутнім навіть на ефективному ринку.

Точна оцінка ступеня прогнозованості волатильності ускладнюється неможливістю спостерігати навіть *ex post* волатильність. Необхідно відзначити, що завдання моделювання волатильності стає все більш і більш важливою для безлічі фінансових додатків, пов'язаних з оцінкою ступеня ризику різних операцій, особливо з причини глобалізації фінансових ринків і постійної появи нових фінансових інструментів, і цілий ряд теоретичних і емпіричних досліджень свідчить про наявність реальної можливості такого роду моделювання. При цьому хотілося б привернути окрему увагу до важливості аналізу прогнозної сили моделей поза вибіркою, так як багато авторів обмежуються виключно статистичними тестами якості підгонки в межах вибірки.

Таким чином, залишається остаточно очевидним, чи існує модель, дійсно дозволяє перевершити "випадкове блукання" в економічному (а не тільки статистичному) сенсі на фінансових ринках взагалі і на валютних ринках зокрема, а також яка точна ступінь прогнозованості волатильності, однак можна однозначно стверджувати, що моделі з довгою пам'яттю безумовно мають потенціал для їх використання в подальших дослідженнях і необхідно продовжувати пошук

можливостей по їх подальшого доопрацювання. Серед найбільш цікавих напрямків роботи можна згадати розширення моделі за рахунок включення в неї макроекономічних змінних (таких, як, наприклад, диференціал процентних ставок між країнами), перехід до багатовимірному нагоди для вивчення взаємного впливу різних валютних ринків, а також використання внутрішньоденних даних для розрахунку реальної волатильності і більш коректної оцінки її прогнозованості.

На закінчення можна сказати, що процес вивчення поведінки фінансових ринків ще далекий від завершення, і значний обсяг роботи залишається поки не здійсненим, проте багато важливих аспектів вже описані і нові дослідження, які проливають світло на функціонування фінансових ринків, безсумнівно підуть в найближчому майбутньому.

ПЕРЕЛІК ПОСИЛАНЬ

1. Andersen T.G. Answering the Critics: Yes, ARCH Models Do Provide Good Volatility Forecasts / Andersen T.G., Bollerslev T. // NBER working paper. – 1997. – N 6023. – P. 110-146.
2. Andersen T.G. Deutsche Mark-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements and Longer Run Dependencies / Andersen T.G., Bollerslev T. // Journal of Finance. – 1998. – N 53(1). – P. 219-265.
3. Andersson M.K. On the Effects of Imposing or Ignoring Long Memory When Forecasting / Andersson M.K. // Working Paper Series in Economics and Finance. – 1998. – N 225. – P. 4-38.
4. Baillie R.T. Long Memory Processes and Fractional Integration in Econometrics // Journal of Econometrics. – 1996. – N 73. – P. 5-59.
5. Baillie R.T. Intra Day and Inter Day Volatility in Foreign Exchange Rates / Baillie R.T., Bollerslev T. // Review of Economic Studies. – 1991. – N 58. – P. 565-585.
6. Baillie R.T. Fractionally integrated generalized autoregressive conditional heteroscedasticity / Baillie R.T., Bollerslev T., Mikkelsen H.O. // Journal of econometrics. – 1996. – N 74. – P. 3-30.
7. Baillie R.T. Analysing inflation by the fractionally integrated ARFIMA — GARCH model / Baillie R.T., Chung C.-F., Tieslau M.A. // Journal of applied econometrics. – 1996. – N 11. – P. 23-40.
8. Persistent dependence in foreign exchange rates? A reexamination / [Barkoulas J.T., Baum C.F., Caglayan M., Chakraborty A.] // Working Paper Boston College. – 2000. – N 377. – P. 3-91.

9. Beechey M. The Efficient Market Hypotheses: A Survey / Beechey M., Gruen D., Vickery J. // Research Discussion Paper, Reserve Bank of Australia. – 2000. – N 203. – P. 4-127.
10. Beine M. La persistance des chocs de volatilité sur le marché des changes s'est-elle modifiée depuis le début des années quatre-vingts? / Beine M., Laurent S. // La Revue Economique. – 2000. – N 3. – P. 703-711.
11. Beine M. Accounting for conditional leptokurtosis and closing days effects in FIGARCH models of daily exchange rates / Beine M., Laurent S., Lecourt C. // Applied Financial Economics. – 2002. – N 12. – P. 589-600.
12. Long-Memory Processes: Probabilistic Properties and Statistical Methods. [Электронный ресурс] / [Jan Beran, Rafal Kulik, Sucharita Ghosh, Yuanhua Feng]. – 2013 Режим доступа: https://www.researchgate.net/publication/279392966_Long-Memory_Processes.

ДОДАТОК А ПРИКЛАД ТЕКСТУ ПРОГРАМИ

```
## ----echo=FALSE,eval=TRUE-----
options(continue=" ")

## -----
options(digits=3)
options(width=72) # narrow output
ds = read.csv("http://www.amherst.edu/~nhorton/r2/datasets/help.csv")
library(dplyr)
newds = select(ds, cesd, female, i1, i2, id, treat, fla, flb, flc, fld, fle,
  flf, flg, flh, fli, flj, flk, fll, flm, fln, flo, flp, flq, flr, fls, flt)

## -----
names(newds)
str(newds[,1:10]) # structure of the first 10 variables

## -----
summary(newds[,1:10]) # summary of the first 10 variables

## -----
head(newds, n=3)

## -----
comment(newds) = "HELP baseline dataset"
comment(newds)
save(ds, file="savedfile")

## -----
write.csv(ds, file="ds.csv")

## -----
library(foreign)
write.foreign(newds, "file.dat", "file.sas", package="SAS")

## -----
with(newds, cesd[1:10])
with(newds, head(cesd, 10))

## -----
with(newds, cesd[cesd > 56])

## -----
library(dplyr)
filter(newds, cesd > 56) %>% select(id, cesd)

## -----
with(newds, sort(cesd)[1:4])
with(newds, which.min(cesd))

## -----
library(mosaic)
tally(~ is.na(flg), data=newds)
favstats(~ flg, data=newds)
```

```

## -----
# reverse code f1d, f1h, f1l and f1p
cesditems = with(newds, cbind(f1a, f1b, f1c, (3 - f1d), f1e, f1f, f1g,
  (3 - f1h), f1i, f1j, f1k, (3 - f1l), f1m, f1n, f1o, (3 - f1p),
  f1q, f1r, f1s, f1t))
nmisscesd = apply(is.na(cesditems), 1, sum)
ncesditems = cesditems
ncesditems[is.na(cesditems)] = 0
newcesd = apply(ncesditems, 1, sum)
imputemeancesd = 20/(20-nmisscesd)*newcesd

## -----
data.frame(newcesd, newds$cesd, nmisscesd, imputemeancesd)[nmisscesd>0,]

## ----createdrink,message=FALSE-----
library(dplyr)
library(memisc)
newds = mutate(newds, drinkstat=
  cases(
    "abstinent" = i1==0,
    "moderate" = (i1>0 & i1<=1 & i2<=3 & female==1) |
      (i1>0 & i1<=2 & i2<=4 & female==0),
    "highrisk" = ((i1>1 | i2>3) & female==1) |
      ((i1>2 | i2>4) & female==0)))

## ----echo=FALSE-----
library(mosaic)

## ----echo=FALSE-----
detach(package:memisc)
detach(package:MASS)

## -----
library(dplyr)
tmpds = select(newds, i1, i2, female, drinkstat)
tmpds[365:370,]

## -----
library(dplyr)
filter(tmpds, drinkstat=="moderate" & female==1)

## ----message=FALSE-----
library(gmodels)
with(tmpds, CrossTable(drinkstat))

## -----
with(tmpds, CrossTable(drinkstat, female,
  prop.t=FALSE, prop.c=FALSE, prop.chisq=FALSE))

## -----
newds = transform(newds,
  gender=factor(female, c(0,1), c("Male","Female")))
tally(~ female + gender, margin=FALSE, data=newds)

## -----
library(dplyr)
newds = arrange(ds, cesd, i1)

```

```

newds[1:5, c("cesd", "i1", "id")]

## -----
library(dplyr)
females = filter(ds, female==1)
with(females, mean(cesd))
# an alternative approach
mean(ds$cesd[ds$female==1])

## -----
with(ds, tapply(cesd, female, mean))
library(mosaic)
mean(cesd ~ female, data=ds)

## ----echo=FALSE,eval=TRUE-----
options(continue=" ")

## -----
x = seq(from=-4, to=4.2, length=100)
normval = dnorm(x, 0, 1)
dfval = 1
tval = dt(x, df=dfval)

## ----normplot-----
plot(x, normval, type="n", ylab="f(x)", las=1)
lines(x, normval, lty=1, lwd=2)
lines(x, tval, lty=2, lwd=2)
legend(1.1, .395, lty=1:2, lwd=2,
      legend=c(expression(N(mu == 0, sigma == 1)),
        paste("t with ", dfval, " df", sep="")))

## ----xpnorm,echo=TRUE, eval=TRUE-----
library(mosaic)
xpnorm(1.96, mean=0, sd=1)

## ----echo=FALSE,eval=TRUE-----
options(continue=" ")

## -----
options(digits=3)
options(width=72) # narrows output to stay in the grey box
ds = read.csv("http://www.amherst.edu/~nhorton/r2/datasets/help.csv")

## -----
with(ds, mean(cesd))

## -----
with(ds, median(cesd))
with(ds, range(cesd))
with(ds, sd(cesd))
with(ds, var(cesd))
library(mosaic)
favstats(~ cesd, data=ds)

## -----

```

```

library(moments)
with(ds, skewness(cesd))
with(ds, kurtosis(cesd))

## -----
with(ds, quantile(cesd, seq(from=0, to=1, length=11)))

## ----bplotr,echo=TRUE,eval=TRUE-----
with(ds, hist(cesd, main="", freq=FALSE))
with(ds, lines(density(cesd), main="CESD", lty=2, lwd=2))
xvals = with(ds, seq(from=min(cesd), to=max(cesd), length=100))
with(ds, lines(xvals, dnorm(xvals, mean(cesd), sd(cesd)), lwd=2))

## ----cormat1-----
cormat = cor(with(ds, cbind(cesd, mcs, pcs)))
cormat

## ----cormat2-----
cormat[c(2, 3), 1]

## ----rugplotr,echo=TRUE,eval=TRUE-----
with(ds, plot(cesd[female==1], mcs[female==1], xlab="CESD", ylab="MCS",
  type="n", bty="n"))
with(ds, text(cesd[female==1 & substance=="alcohol"],
  mcs[female==1 & substance=="alcohol"], "A"))
with(ds, text(cesd[female==1 & substance=="cocaine"],
  mcs[female==1 & substance=="cocaine"], "C"))
with(ds, text(cesd[female==1 & substance=="heroin"],
  mcs[female==1 & substance=="heroin"], "H"))
with(ds, rug(jitter(mcs[female==1]), side=2))
with(ds, rug(jitter(cesd[female==1]), side=3))

## ----message=FALSE-----
require(gmodels)
with(ds, CrossTable(homeless, female, prop.chisq=FALSE, format="SPSS"))

## -----
or = with(ds, (sum(homeless==0 & female==0)*
  sum(homeless==1 & female==1))/
  (sum(homeless==0 & female==1)*
  sum(homeless==1 & female==0)))
or

## -----
library(epitools)
oddsobject = with(ds, oddsratio.wald(homeless, female))
oddsobject$measure
oddsobject$p.value

## -----
chisqval = with(ds, chisq.test(homeless, female, correct=FALSE))
chisqval

## -----
with(ds, fisher.test(homeless, female))

## ----gridtable,echo=TRUE, eval=TRUE,message=FALSE-----

```

```

library(gridExtra)
mytab = tally(~ racegrp + substance, data=ds)
plot.new()
grid.table(mytab)

## -----
ttres = t.test(age ~ female, data=ds)
print(ttres)

## ----message=FALSE-----
library(coin)
oneway_test(age ~ as.factor(female),
  distribution=approximate(B=9999), data=ds)

## -----
with(ds, wilcox.test(age ~ as.factor(female), correct=FALSE))

## ----kstest-----
ksres = with(ds, ks.test(age[female==1], age[female==0]))
print(ksres)

## -----
plotdens = function(x,y, mytitle, mylab) {
  densx = density(x)
  densy = density(y)
  plot(densx, main=mytitle, lwd=3, xlab=mylab, bty="l")
  lines(densy, lty=2, col=2, lwd=3)
  xvals = c(densx$x, rev(densy$x))
  yvals = c(densx$y, rev(densy$y))
  polygon(xvals, yvals, col="gray")
}

## ----polyplotr,echo=TRUE,eval=TRUE-----
mytitle = paste("Test of ages: D=", round(ksres$statistic, 3),
  " p=", round(ksres$p.value, 2), sep="")
with(ds, plotdens(age[female==1], age[female==0], mytitle=mytitle,
  mylab="age (in years)"))
legend(50, .05, legend=c("Women", "Men"), col=1:2, lty=1:2, lwd=2)

## -----
library(survival)
survobj = survdiff(Surv(dayslink, linkstatus) ~ treat,
  data=ds)
print(survobj)
names(survobj)

## ----echo=FALSE,eval=TRUE-----
options(continue=" ")

## -----
options(digits=3)
cat = as.factor(rep(c("blue", "red"), each=50))
y = rnorm(100)
tapply(y, cat, mean)

## -----

```

```

xtabs(~ ave(y, cat, FUN=mean))

## -----
library(mosaic)
mean(y ~ cat)

## -----
library(dplyr)
ds = data.frame(y, cat)
groups = group_by(ds, cat)
summarise(groups, mean=mean(y))

## -----
ds = read.csv("http://www.amherst.edu/~nhorton/r2/datasets/help.csv")
uniquevals = unique(ds$female)
numunique = length(uniquevals)
formula = as.formula(i1 ~ age)
p = length(coef(lm(formula, data=ds)))
params = matrix(rep(0, numunique*p), nrow=p, ncol=numunique)
for (i in 1:length(uniquevals)) {
  cat("grouping:", i, "\n")
  params[i,] = coef(lm(formula, data=subset(ds,
    female==uniquevals[i])))
}
rownames(params) = c("Intercept", "Age")
colnames(params) = ifelse(uniquevals==0, "male", "female")
params

## ----message=FALSE-----
library(plyr)
models = dlply(ds, "female", function(df) {
  lm(i1 ~ age, data=df)
})
ldply(models, coef)

## ----message=FALSE-----
library(MASS)
library(nlme)
# initialize parameters and building blocks
effect = 0.35 # effect size
corr = 0.4    # intrafamilial correlation
numsim = 1000
n1fam = 50    # families with 3 exposed
n2fam = 50    # families with 1 exposed and 2 unexposed
# 3x3 compound symmetry matrix
vmat = matrix(c
  ( 1,  corr, corr,
    corr, 1  , corr,
    corr, corr, 1  ), nrow=3, ncol=3)
# 1 1 1 ... 1 0 0 0 ... 0
x = c(rep(1, n1fam), rep(1, n1fam), rep(1, n1fam),
      rep(1, n2fam), rep(0, n2fam), rep(0, n2fam))
# 1 2 ... n1fam 1 2 ... n1fam ...
id = c(1:n1fam, 1:n1fam, 1:n1fam,
      (n1fam+1:n2fam), (n1fam+1:n2fam), (n1fam+1:n2fam))
power = rep(0, numsim) # initialize vector for results

```

```

## -----
for (i in 1:numsim) {
  # all three exposed
  grp1 = mvnrm(n1fam, c(effect, effect, effect), vmat)

  # only first exposed
  grp2 = mvnrm(n2fam, c(effect, 0, 0), vmat)

  # concatenate the output vector
  y = c(grp1[,1], grp1[,2], grp1[,3],
        grp2[,1], grp2[,2], grp2[,3])

  group = groupedData(y ~ x | id) # specify dependence structure
  res = lme(group, random = ~ 1) # fit random intercept model
  pval = summary(res)$tTable[2,5] # grab results for main parameter
  power[i] = pval<=0.05 # is it statistically significant?
}

## -----
cat("\nEmpirical power for effect size of ", effect,
    " is ", round(sum(power)/numsim,3), ".\n", sep="")
cat("95% confidence interval is",
    round(prop.test(sum(power), numsim)$conf.int, 3), "\n")

## ----eval=FALSE-----
## library(markdown); library(knitr)
## knit("filename.Rmd") # creates filename.md
## markdownToHTML("filename.md", "filename.html")
## browseURL("filename.html")

## ----eval=FALSE-----
## library(knitr)
## knit("filename.Rnw")

## ----eval=FALSE-----
## library(MCMCpack)
## # linear regression
## mod1 = MCMCregress(formula, burnin=1000, mcmc=10000, data=ds)
## # logistic regression
## mod2 = MCMClogit(formula, burnin=1000, mcmc=10000, data=ds)
## # Poisson regression
## mod3 = MCMCpoisson(formula, burnin=1000, mcmc=10000, data=ds)

## ----message=FALSE-----
events.0=0 # for X = 0
events.1=5 # for X = 1
x = as.factor(c(rep(0,100), rep(1,100)))
y = c(rep(0,100-events.0), rep(1,events.0),
      rep(0, 100-events.1), rep(1, events.1))

library(MCMCpack)
logmcmc = MCMClogit(y ~ x, burnin=100, mcmc=2000, b0=0, B0=.04)

## -----
summary(logmcmc)

## -----

```



```

library(MCMCpack)
posterior = with(ds,
  MCMCpoisson(i1 ~ female + as.factor(substance) + age,
    burnin=100, mcmc=2000))

## -----
summary(posterior)

## -----
lm1 = lm(pcs ~ homeless, data=ds)
summary(lm1)

## -----
lm2 = lm(pcs ~ homeless + age + female + i1 + mcs, data=ds)
summary(lm2)

## -----
form = formula(homeless ~ age + female + i1 + mcs)
glm1 = glm(form, family=binomial, data=ds)
propensity = glm1$fitted

## -----
lm3 = lm(pcs ~ homeless + propensity, data=ds)
summary(lm3)

## ----message=FALSE-----
library(Matching)
rr = with(ds, Match(Y=pcs, Tr=homeless, X=propensity, M=1))
summary(rr)

## ----capture-----
longout = capture.output(MatchBalance(form, match.out=rr,
  nboots=10, data=ds))
write(longout[1:20], file = "")

## ----echo=TRUE,eval=TRUE-----
x = rnorm(1000, mean=1)

## ----echo=TRUE,eval=TRUE-----
covfun = function(x) { # multiply CV by 100
  return(100*sd(x)/mean(x))
}

## ----message=FALSE,echo=TRUE,eval=TRUE-----
options(digits=4)
res2 = replicate(2000, covfun(sample(x, replace=TRUE)))
quantile(res2, c(.025, .975))

## ----message=FALSE,echo=TRUE,eval=TRUE-----
options(digits=4)
covfun(x)
library(mosaic)
res = do(2000) * covfun(resample(x))
quantile(res$result, c(.025, .975))

## -----
x = c(1, 2, NA)

```

```

mean(x)

## -----
mean(x, na.rm=TRUE)

## -----
sum(na.omit(x))

## -----
sum(!is.na(x))

## -----
# remap values of x with missing value code of 999 to missing
w = c(1, 2, 999)
w[w==999] = NA
w

## -----
w = c(1,2,999)

is.na(w) = w==999 # set 999's to missing
w

## ----misshelp-----
ds =
  read.csv("http://www.amherst.edu/~nhorton/r2/datasets/helpmiss.csv")
smallds = with(ds, data.frame(homeless, female, i1, sexrisk, indtot,
  mcs, pcs))

## ----echo=FALSE-----
options(width=72)

## ----message=FALSE-----
summary(smallds)
library(Hmisc)
na.pattern(smallds)

## -----
glm(homeless ~ female + i1 + sexrisk + indtot, binomial,
  data=smallds)

## ----message=FALSE-----
library(mice)
imp = mice(smallds, m=20, maxit=25, seed=42, print=FALSE)

## -----
summary(pool(glm.mids(homeless ~ female + i1 + sexrisk +
  indtot, family=binomial, data=imp)))

## -----
set.seed(1492)
n = 10000
x = rnorm(n)
probgroupl = exp(-1 + 2*x)/(1 + exp(-1 + 2*x))
group = ifelse(probgroupl > runif(n), 1, 0)
y = (group * 3) + ((1 + group/5) * x) + rnorm(n)

```

```

## -----
library(flexmix)
mixout.fm=flexmix(y ~ x, k=2, model=FLXMRglmfix(y ~ x, varFix=TRUE),
  concomitant=FLXPmultinom(~ x))

## -----
parameters(mixout.fm)

## -----
parameters(mixout.fm, which="concomitant")
## ----echo=FALSE,eval=TRUE-----
options(continue=" ")

## -----
options(digits=3)
cat = as.factor(rep(c("blue","red"), each=50))
y = rnorm(100)
tapply(y, cat, mean)

## -----
xtabs(~ ave(y, cat, FUN=mean))

## -----
library(mosaic)
mean(y ~ cat)

## -----
library(dplyr)
ds = data.frame(y, cat)
groups = group_by(ds, cat)
summarise(groups, mean=mean(y))

## -----
ds = read.csv("http://www.amherst.edu/~nhorton/r2/datasets/help.csv")
uniquevals = unique(ds$female)
numunique = length(uniquevals)
formula = as.formula(i1 ~ age)
p = length(coef(lm(formula, data=ds)))
params = matrix(rep(0, numunique*p), nrow=p, ncol=numunique)
for (i in 1:length(uniquevals)) {
  cat("grouping:", i, "\n")
  params[,i] = coef(lm(formula, data=subset(ds,
    female==uniquevals[i])))
}
rownames(params) = c("Intercept", "Age")
colnames(params) = ifelse(uniquevals==0, "male", "female")
params

## ----message=FALSE-----
library(plyr)
models = dlply(ds, "female", function(df) {
  lm(i1 ~ age, data=df)
})
ldply(models, coef)

## ----message=FALSE-----
library(MASS)

```

```

library(nlme)
# initialize parameters and building blocks
effect = 0.35 # effect size
corr = 0.4    # intrafamilial correlation
numsim = 1000
n1fam = 50    # families with 3 exposed
n2fam = 50    # families with 1 exposed and 2 unexposed
# 3x3 compound symmetry matrix
vmat = matrix(c
  ( 1,  corr, corr,
    corr, 1  , corr,
    corr, corr, 1  ), nrow=3, ncol=3)
# 1 1 1 ... 1 0 0 0 ... 0
x = c(rep(1, n1fam), rep(1, n1fam), rep(1, n1fam),
      rep(1, n2fam), rep(0, n2fam), rep(0, n2fam))
# 1 2 ... n1fam 1 2 ... n1fam ...
id = c(1:n1fam, 1:n1fam, 1:n1fam,
      (n1fam+1:n2fam), (n1fam+1:n2fam), (n1fam+1:n2fam))
power = rep(0, numsim) # initialize vector for results

## -----
for (i in 1:numsim) {
  # all three exposed
  grp1 = mvrnorm(n1fam, c(effect, effect, effect), vmat)

  # only first exposed
  grp2 = mvrnorm(n2fam, c(effect, 0, 0), vmat)

  # concatenate the output vector
  y = c(grp1[,1], grp1[,2], grp1[,3],
        grp2[,1], grp2[,2], grp2[,3])

  group = groupedData(y ~ x | id) # specify dependence structure
  res = lme(group, random = ~ 1) # fit random intercept model
  pval = summary(res)$tTable[2,5] # grab results for main parameter
  power[i] = pval<=0.05          # is it statistically significant?
}

## -----
cat("\nEmpirical power for effect size of ", effect,
    " is ", round(sum(power)/numsim,3), ".\n", sep="")
cat("95% confidence interval is",
    round(prop.test(sum(power), numsim)$conf.int, 3), "\n")

## ----eval=FALSE-----
## library(markdown); library(knitr)
## knit("filename.Rmd") # creates filename.md
## markdownToHTML("filename.md", "filename.html")
## browseURL("filename.html")

## ----eval=FALSE-----
## library(knitr)
## knit("filename.Rnw")

## ----eval=FALSE-----
## library(MCMCpack)
## # linear regression

```

```

## mod1 = MCMCregress(formula, burnin=1000, mcmc=10000, data=ds)
## # logistic regression
## mod2 = MCMClogit(formula, burnin=1000, mcmc=10000, data=ds)
## # Poisson regression
## mod3 = MCMCpoisson(formula, burnin=1000, mcmc=10000, data=ds)

## ----message=FALSE-----
events.0=0 # for X = 0
events.1=5 # for X = 1
x = as.factor(c(rep(0,100), rep(1,100)))
y = c(rep(0,100-events.0), rep(1,events.0),
      rep(0, 100-events.1), rep(1, events.1))

library(MCMCpack)
logmcmc = MCMClogit(y ~ x, burnin=100, mcmc=2000, b0=0, B0=.04)

## -----
summary(logmcmc)

## -----
library(MCMCpack)
posterior = with(ds,
  MCMCpoisson(i1 ~ female + as.factor(substance) + age,
    burnin=100, mcmc=2000))

## -----
summary(posterior)

## -----
lm1 = lm(pcs ~ homeless, data=ds)
summary(lm1)

## -----
lm2 = lm(pcs ~ homeless + age + female + i1 + mcs, data=ds)
summary(lm2)

## -----
form = formula(homeless ~ age + female + i1 + mcs)
glm1 = glm(form, family=binomial, data=ds)
propensity = glm1$fitted

## -----
lm3 = lm(pcs ~ homeless + propensity, data=ds)
summary(lm3)

## ----message=FALSE-----
library(Matching)
rr = with(ds, Match(Y=pcs, Tr=homeless, X=propensity, M=1))
summary(rr)

## ----capture-----
longout = capture.output(MatchBalance(form, match.out=rr,
  nboots=10, data=ds))
write(longout[1:20], file = "")

## ----echo=TRUE,eval=TRUE-----
x = rnorm(1000, mean=1)

```

```

## ----echo=TRUE,eval=TRUE-----
covfun = function(x) { # multiply CV by 100
  return(100*sd(x)/mean(x))
}

## ----message=FALSE,echo=TRUE,eval=TRUE-----
options(digits=4)
res2 = replicate(2000, covfun(sample(x, replace=TRUE)))
quantile(res2, c(.025, .975))

## ----message=FALSE,echo=TRUE,eval=TRUE-----
options(digits=4)
covfun(x)
library(mosaic)
res = do(2000) * covfun(resample(x))
quantile(res$result, c(.025, .975))

## -----
x = c(1, 2, NA)
mean(x)

## -----
mean(x, na.rm=TRUE)

## -----
sum(na.omit(x))

## -----
sum(!is.na(x))

## -----
# remap values of x with missing value code of 999 to missing
w = c(1, 2, 999)
w[w==999] = NA
w

## -----
w = c(1,2,999)

is.na(w) = w==999 # set 999's to missing
w

## ----misshelp-----
ds =
  read.csv("http://www.amherst.edu/~nhorton/r2/datasets/helpmiss.csv")
smalllds = with(ds, data.frame(homeless, female, i1, sexrisk, indtot,
  mcs, pcs))

## ----echo=FALSE-----
options(width=72)

## ----message=FALSE-----
summary(smalllds)
library(Hmisc)
na.pattern(smalllds)

```

```

## -----
glm(homeless ~ female + i1 + sexrisk + indtot, binomial,
    data=smallds)

## ----message=FALSE-----
library(mice)
imp = mice(smallds, m=20, maxit=25, seed=42, print=FALSE)

## -----
summary(pool(glm.mids(homeless ~ female + i1 + sexrisk +
    indtot, family=binomial, data=imp)))

## -----
set.seed(1492)
n = 10000
x = rnorm(n)
probgroupl = exp(-1 + 2*x)/(1 + exp(-1 + 2*x))
group = ifelse(probgroupl > runif(n), 1, 0)
y = (group * 3) + ((1 + group/5) * x) + rnorm(n)

## -----
library(flexmix)
mixout.fm=flexmix(y ~ x, k=2, model=FLXMRglmfix(y ~ x, varFix=TRUE),
    concomitant=FLXPmultinom(~ x))

## -----
parameters(mixout.fm)

## -----
parameters(mixout.fm, which="concomitant")

```